A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

Three Situational Determinants of Conformity to Social Norms 325
Solomon C. Goldberg

Political Attitudes and Judgments of Other People 330
Gustav Jahoda

Reaction Time as a Function of Manifest Anxiety and Stimulus Intensity 335
Charles Wenar

Some Effects of Cooperation and Competition upon Small Group Behavior 341
Morton M. Grossack

Ethnocentrism and Tolerance of Trait "Inconsistency" 349
Ivan D. Steiner

The Effect of Stuttering on the Behavior of the Listener 355
Seymour Rosenberg and James Curtiss

Complexity of Response as a Factor in the Vocabulary Performance of Schizophrenics 362
R. Herrington and J. C. Ehrmann

Authoritarianism and Leadership Choice in a Military Setting 365
E. P. Hollander

Test Anxiety and Rorschach Performance 371
Francis N. Cox and Seymour B. Sarason

Religious Beliefs of Catholic College Students and Their Attitudes toward Minorities 378
Charles T. O'Reilly and Edward J. O'Reilly

Continued on inside front cover

Volume 49 JULY 1954 Number

CONTENTS—*Continued*

ARTICLES PAGE

Assumed Similarity Measures as Predictors of Team Effectiveness 381
Fred E. Fiedler

The Effect of Reward on Adult Imitative Behavior 389
Edgar H. Schein

Is There a Mechanism of Perceptual Defense? A Reply to Postman, Bronson, and Gropper 396
Richard S. Lazarus

A Theoretical Approach to Psychological Movement 399
Alvin R. Howard and George A. Kelly

The Effect of Experimental Arousal of the Affiliation Motive on Thematic Apperception 405
John W. Atkinson, Roger W. Heyns, and Joseph Veroff

Interpersonal Knowledge and Individual and Group Effectiveness 411
F. Loyal Gage, Eugene H. Colorez, and Peter D. Wardlow

Irregularities in Judgment Data Collected by the Method of Equal-Appearing Intervals 415
Sam C. Webb

Perceptual Rigidity as Measured by Aniseikonic Lenses 419
Wesley C. Becker

Values as Determinants of Word-Recognition Thresholds 423
J. C. Gilchrist, J. F. Lederman, and William Lee

Word Frequency and the Measurement of Value Areas 427
Donald R. Brown and Joe Adams

The Assessment of Group Opinion by Leaders, and Their Influence on Its Formation 431
George A. Talland

Prediction in the Clinical Method and Interpretations of Biochemistry, Psychiatry, and Psychology 435
Eli Robins and Leon N. Mandi

Affective Tone and Visual Recognition Thresholds 443
Leonard D. Goodstein

Some Effects of Situational Threat on Group Behavior 445
John T. Lanzetta, Don Haefner, Peter Langham, and Howard Axelrod

A Study in Language and Cognition 454
Roger W. Brown and Eric H. Lenneberg

CRITIQUE AND NOTES

Some Psychodynamic Aspects of Felt Parental Affiliation in Young Children 463
Edward M. Bennett and Dorothea K. Zakarewski

Autokinetic Judgments as a Measure of Influence 464
Harriet B. Linton

Social Research and the Problem of Values: A Reply 466
Else Frenkel-Brunswik

Administration of the F Scale to a Sample of Germans 471
Thomas S. Cohn and Henry Cowen

Notes on the Wesley Rigidity Scale: The Development of a Short Form 472
Seymour L. Zelen and Eugene H. Levin

REVIEWS 474

NEW BOOKS RECEIVED 481

Volume 49 JULY 1954 Number 3

The Journal of Abnormal and Social Psychology

Editor: J. McV. Hunt, University of Illinois
Associate Editor: Arthur Jenness, Williams College
Book Review Editor: Stanley G. Estes, Harvard University
Managing Editor: Lorraine Bouthilet

CONSULTING EDITORS

Gordon W. Allport
Harvard University
Hadley Cantril
Princeton University
Lee J. Cronbach
University of Illinois
John Dollard
Yale University
Leonard W. Doob
Yale University
Leon Festinger
University of Minnesota
Charles M. Harsh
U.S. Navy Electronics Laboratory
San Diego
William A. Hunt
Northwestern University
W. Ernst Kris
Yale University School of Medicine
Donald W. MacKinnon
University of California
Quinn McNemar
Stanford University
O. Hobart Mowrer
University of Illinois
Lawrence I. O'Kelly
University of Illinois
Charles E. Osgood
University of Illinois
Saul Rosenzweig
Washington University
Harold Seashore
Psychological Corporation
David Shakow
University of Illinois School of Medicine
Richard L. Solomon
Harvard University
Ross Stagner
University of Illinois
Ivan Steiner
University of Illinois
Elizabeth Collins Wesman
White Plains, N. Y.
Robert W. White
Harvard University
Charles Wrigley
University of Illinois

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.
AND MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1954 by the American Psychological Association, Inc.

The JOURNAL is issued quarterly in January, April, July, and October. The yearly volume averages 544 pages. The subscription price per volume is $7.00; foreign, $7.50. Single issues are $2.00.

Original manuscripts and editorial communications should be sent to:

DR. J. MCV. HUNT
Department of Psychology
University of Illinois
Urbana, Ill.

Books for review should be sent to:

DR. STANLEY G. ESTES
Psychological Clinic
64 Plympton Street
Cambridge 38, Mass.

Fifty free offprints are supplied to contributors of articles. For book reviews, five copies of the JOURNAL are supplied gratis if requested. Authors of prior publication articles receive no gratis offprints or journals.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

Subscriptions and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

THREE SITUATIONAL DETERMINANTS OF CONFORMITY TO SOCIAL NORMS

SOLOMON C. GOLDBERG

Army Medical Service Graduate School, Walter Reed Army Medical Center[1]

GROUP psychology has been variously depicted as the establishment of social norms by the group and the adoption of these norms by the members of the group. The present study is concerned with certain of the conditions under which group members will conform to the social norms of their groups to one degree or another. We shall consider primarily three variables which, as yet, have received little systematic attention. These are: (*a*) the number of occasions a subject (*S*) has been exposed to the social norm (to be known as "exposures"), (*b*) the extent of *S*'s agreement or disagreement with the social norm when he is initially exposed to it (to be known as "distance"), and (*c*) the size of the group to which the norm is attributed.

The importance of exposures seems to be quite well established in advertising practices; it is assumed that the consumer tends to adopt a more favorable attitude toward the advertised product the more often he is told of the vast number of people and prestige individuals who already use the product.

The distance variable in relation to conformity has been observed in a number of experiments but has not been systematically studied. The best example is Sherif's autokinetic experiment (4), which showed that the individuals who initially held the more extreme opinions were those who exhibited the greatest opinion change. In another setting Schachter (3) found that the group as a whole directed more communications toward the group member whose opinion deviated most from the group norm opinion in an attempt to bring him more in line with the group. From these results it would follow that it will be the opinion-deviate who will conform most to the group, i.e., change his attitude most in favor of the group attitude.

[1] This paper represents part of a doctoral dissertation completed in 1952 at the Catholic University of America under the supervision of Drs. P. R. Hofstaetter, J. W. Stafford, and M. Lorr.

The effect of group size on conformity has been explored by Asch (1). He found that larger groups, up to a given size, exert greater influence upon the group member whose opinion differs from the group's. The present experiment presents the opportunity to test the differential effects on conformity of groups of two and groups of four.

HYPOTHESES

The preceding introductory discussion may be summarized in terms of the hypotheses to be tested in the present experiment. These are that conformity is greater (*a*) when the number of exposures to the group norm is greater, (*b*) when the group member's initial distance from the group norm is greater, and (*c*) when the size of the group is larger.

METHOD

Subjects

The *S*s were 79 white males who were either college or high school students or young business men. The sample was subdivided into nine experimental groups of four each, nine experimental groups of two each, two experimental groups of three each, and one control group of 19 *S*s.

Materials

The experimental stimuli were nine bust-size photographs of male Negroes taken from a high school yearbook. The nine photographs (on the basis of previous judgments by 24 judges) were selected such that the variabilities of their adjudged intelligence were high, and the means and variabilities of their adjudged intelligence did not differ significantly from each other. This procedure allowed the selection of nine equated stimuli which were relatively unstereotyped with regard to adjudged intelligence.

Procedure

Control group. The following instructions were given to the group:

"The purpose of this experiment is to settle once and for all whether people have accuracy in making snap judgments of intelligence from photographs. You all have seen some people who seem to be able to size up other people rather quickly and rather well; so probably some of you will be more accurate than others. I am going to show you nine different pictures. You will see

each picture more than once, but you will see some pictures more than others. The reason for this is that we are trying to see if you become more accurate on those pictures which you have had more chances to see. I will show you each picture for three seconds. During that time look at it carefully and judge the intelligence of that person on a scale from 50, low, to 150, high, depending on how high or how low you think he is. Write down your judgment for each picture as it comes up. It is only fair to tell you that, in order to simplify the experiment, we chose to photograph mainly rather smart and rather dumb people. Although most of these people were either smart or dumb, we could not avoid a very few who were average in intelligence. Try to improve your accuracy every time you see each picture. Any questions?"

The first three photographs were judged 20 times each; the second three, 10 times each; and the last three, 5 times each; thus the total number of judgments was 105. The presentation of the photographs was randomized, and the control group underwent two such judgment sessions separated by a 15-minute rest period.

Experimental groups. Each experimental group was at first given the same instructions as the control group and then completed one session of 105 judgments. A second session of judgments was then administered to each group member individually. This session was preceded by the following instructions:

"In this session you are asked to judge the same pictures you did before. I have looked over your first judgments and have made a note as to what the group as a whole thought of each picture. Now every time I show you a picture, as I did before, I will tell you what judgment the group as a whole made for that picture. Then, as before, look at the picture carefully, make your own judgment on it, and write it down on the sheet. The reason for this is that we want to see if your judgment accuracy increases or decreases when you know what the group judgment is. In other words, will your knowledge of the group judgment aid or hinder your accuracy? Remember, try to improve your judgment accuracy every time you see each picture. Any questions?"

The group estimate of each photograph reported to *S*s in this session was a fictitious one; it was manipulated to be a desired distance from *S*'s estimate of that photograph in the first judgment session. For example, on the first of the nine pictures it was desired that all *S*s' estimates be 50 points from the group estimate; to achieve this the group estimate reported to each *S* for that photograph was 50 points from *S*'s judgment of that photograph in the first session of judgments. More specifically, if in the first judgment session subject A judged the first photograph to be 60 and subject B judged it to be 120, then the fictitious group opinions given to them were 110 and 70 respectively. Whenever possible, the fictitious group opinion was reported in the direction opposed to the midpoint value of 100. For example, with a distance of 10, if *S*'s opinion were 75, the possible fictitious group opinions could be 65 or 85. In this case, 65 would be preferred in order to counteract any chance regression to the midpoint. An outline of the experimental conditions may be given as follows:

Exposures:	20	20	20	10	10	10	5	5	5
Photograph:	A	B	C	D	E	F	G	H	I
Distance:	50	30	10	50	30	10	50	30	10

This means that on Photograph E, for example, all *S*s were 30 points from the group, and that the fictitious group opinion was exposed to each *S* each of the 10 times he was asked to judge this photograph. These 10 exposures were randomly dispersed with the presentation of the other eight photographs.

RESULTS

For the experiment to be successful it was necessary that *S*s be unaware of its true purpose lest they become resistant to its efforts. At the end of the experimental procedure each *S* was questioned in order to ascertain whether he had "caught on" to our true purpose. It was found that only 25 per cent of the experimental *S*s thought that attempts were being made to influence their opinions. This split of percentages yields a highly significant critical ratio of 6.00, signifying that the procedure had achieved its desired end.

The conformity score. The amount of change in rating points on any given photograph made by an *S*, from his last judgment in the first session to his last judgment in the second session, was taken as his conformity score. For the experimental *S*s the direction of change had to be taken into account. If the change was in the direction of the fictitious group opinion, it was given a positive sign; if away from the fictitious group opinion, it was denoted as negative change. No change in opinion was denoted as zero conformity. Direction of change for control *S*s from the first to the second judgment session was randomly assigned since no fictitious group opinion was reported to them.

Comparison of experimental and control groups. To test for a significant difference between conformity in the experimental groups and opinion change in the control group a *t* test was computed. Its value of 3.59 was significant beyond the .01 level of confidence showing that the experimental manipulations produced a real conforming effect. (See Table 1.)

Conformity to "groups" of four and "groups" of two. The question asked of this analysis is whether an *S* conforms more to a group opinion when he knows the group is composed of a greater number of members even though he is told of the group opinion by a single informant. Certainly in extreme cases of group size this relationship may be expected to hold; however, it appears that in the present limited case it does not. Table 1 (second row) shows a

TABLE 1

CONFORMITY DIFFERENCES: EXPERIMENTAL VS. CONTROL GROUPS; GROUPS OF FOUR VS. GROUPS OF TWO

COMPARISONS	MEAN	σ^2	N	t	p
Experimental	9.5	74.5	60		
vs.				3.59	<.01
Control	1.6	47.7	19		
Groups of four	10.2	73.7	39		
vs.				0.78	>.05
Groups of two	8.3	73.8	21		

Note: Data on two groups of three were not meaningfully analyzable. These two groups were divided, the data of one included in the groups of four and the data of the other in the groups of two.

t of .78, which is not significant to the .05 level of confidence.

Conformity as a function of distance and exposures. The conformity scores for each experimental condition were arranged in a three-by-three factorial design so that an analysis of variance could test for significant differences in conformity between the three conditions of distance from the group norm and between the three conditions of exposures to the group norm; in addition, this analysis would reveal any significant interaction between distance and exposures.

Table 3 shows that the F test between distances is significant beyond the .001 level of confidence; the F's for between exposures and the interaction of distances and exposures are not significant to the .05 level of confidence. The t tests showed no significant differences between conformity at distance 50 and distance 30. However, differences in conformity between distances 30 and 10 and between distances 50 and 10 were significant.

Proportional conformity. The preceding analysis shows the degree of conformity to be expected under a variety of distance conditions. However, the absolute extent of conformity cannot be generalized to situations beyond the present experiment. Of more generality would be a percentage conformity score. This index is taken as the conformity score divided by the distance; in simpler terms it denotes how much an S conformed in terms of how much he was "asked to conform." If this index can be shown to be constant for all conditions of distance, then it might be applied to other situations in which we are trying to predict the degree of opinion change from a knowledge of opinion distance.

TABLE 2

CONFORMITY AS A FUNCTION OF DISTANCE AND EXPOSURES

(N = 540; m per cell = 60)

No. of EXPOSURES		DISTANCE		
		50	30	10
20	M	14.5	11.0	.8
	σ^2	359.15	258.40	198.95
10	M	14.1	11.2	3.3
	σ^2	440.85	301.12	165.20
5	M	15.2	12.5	3.1
	σ^2	470.96	327.95	169.87

TABLE 3

ANALYSIS OF VARIANCE OF CONFORMITY AS A FUNCTION OF DISTANCE AND EXPOSURES

SOURCE	df	VARIANCE	F	p
Distances	2	7283	24.80	<.001
Exposures	2	105	<1.00	>.05
Interaction	4	39	<1.00	>.05
Within	531	294		

To test this possibility, all the conformity scores were transformed into proportional conformity scores; an analysis of variance was computed to test for differences between proportional conformity under the nine experimental conditions. The resulting F value of .76 shown in Table 4 was not significant to the .05 level. Thus, we cannot reject the hypothesis that there are no real differences in proportional conformity under varying distance conditions. The index of proportional conformity appears to be a constant whose average value is approximately .30. This means that, within the limits of the present experiment, *S*s reduce their disagreement with the group by approximately 30 per cent regardless of the extent of their initial disagreement.

Conformity as multidimensional behavior. Since we have nine somewhat different social situations in the experimental conditions, we may ask to what extent an individual's conformity in one condition is related to his conformity in another condition. To investigate

TABLE 4

ANALYSIS OF VARIANCE OF PROPORTIONAL CONFORMITY AS A FUNCTION OF DISTANCE

SOURCE	*df*	VARIANCE	*F*	*p*
Distances	8	0.57	0.76	>.05
Within	531	0.75		

Note: Means and variances for each experimental condition can be computed directly from Table 2.

this possibility, intercorrelations[2] were computed between conformity scores in each of the experimental conditions. These ranged in value from −.24 to .47, with the median at .14. It would appear that there is little tendency for individuals to conform to the same extent from one situation to another.

DISCUSSION

The results concerning group size were negative; *S*s did not conform more to the opinions of groups of four than to groups of two. This may have occurred for two possible reasons. First, the difference in group size may not have been sufficient to elicit any increased conformity. Second, since the effect of the group opinion upon the individual was indirect (reported to *S* by *E*), *S*s could have been conforming to *E* rather than to the group; hence, differences in group size would make no difference in conformity. Whichever interpretation is the true one, it was still possible to measure the conditions of conformity differentials to a prestige source, whether this source was the group or *E*.

The most clear-cut results were those concerning the distance variable. A priori, there is no reason to expect *S*s to conform more when the distance is greater. Schachter indicated that the opinion-deviate conformed more because the group exerted greater pressure upon him to conform by directing more communications to him. In the present experiment we see that those with more deviating opinions conformed more, but the pressure brought to bear by the prestige source upon the different degrees of deviation was constant. Schachter's interpretation would not explain the present results. It would appear instead that *S* perceives his distance from the group as an undesirable deviation and then makes a compromise with the group. Perhaps greater deviations are perceived as requiring greater compromises. It is also very interesting that the percentage of distance compromised is a constant for all degrees of distance. Why this is so defies explanation by the available data. The next step in this research should be the investigation of introspectional correlates of conformity at various degrees of distance in an attempt to understand the dynamics underlying these empirical results.

Consider now the number of exposures to the group norm. It was shown that conformity occurred within the first five exposures to the group norm and that additional exposures were not accompanied by additional conformity. This finding concerning the effects of repetition is certainly not in line with the current assumptions of advertising and propaganda. The results do not appear to characterize a situation wherein an *S*'s opinion is gradually weakened by the influence of the group. It would be our guess that *S*s perceive the group opinion as an additional datum upon which to base a new judgment of the stimulus. This new information is taken into account, and, as we have seen, *S* makes some sort of compromise adjustment in the direction of the group. Additional repetitions of the same group opinion over and over do not provide *S* with any new information. What may be happening in the repetition of advertisements is that more and more people are reached rather than that the same people are influenced to a greater and greater extent. Here, too, we expect that phenomenal data would be revealing.

The low intercorrelations between conformity in the nine experimental conditions provide evidence against the existence of a general personality characteristic of suggesti-

[2] A complete centroid factor analysis of the intercorrelation matrix resulted in three factors; these were rotated "blindly" to simple structure keeping the axes orthogonal. Identification of the factors, however, was far from clear, and consequently it was decided not to burden the report with the tabular details of the analysis. At most, one could identify very indirectly two of the factors as "conformity to larger distances" and "conformity to smaller distances." The intercorrelation matrix and completed factor analysis have been deposited with the ADI. Order Document No. 3989 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25 D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

bility or conformity. The existence of such a general personality characteristic would require the rank ordering of individuals in the different experimental conditions to be more highly related than they were. It would appear that investigations of the acquisition of social norms would more fruitfully concentrate upon the analysis of the immediate situation rather than on stable personality characteristics of *S*.

SUMMARY AND CONCLUSIONS

The present study has been concerned with the determinants of conformity to group norms. The variables under investigation were: (*a*) the extent of initial disagreement between the individual's and the group's opinion (known as "distance"); (*b*) the number of occasions the individual was exposed to the group norm (known as "exposures"); and (*c*) the size of the group. The following results were obtained:

1. Knowledge of a group norm by a subject results in the subject's conforming to (reducing his disagreement with) the group norm.

2. The degree of conformity to groups of four is not different from the degree of conformity to groups of two.

3. The absolute degree of conformity to the group norm is a function of the distance from (the extent of disagreement with) the group norm. In general, greater distance makes for greater conformity.

4. Subjects conform to the group norm by a constant percentage of the distance from the group norm.

5. Conformity occurs within the first few exposures to the group norm; additional exposures to the group norm are not accompanied by additional conformity to the group norm.

6. Intercorrelations between conformity under all experimental conditions were low. Conformity was not generalized but appeared to be highly situational. This was considered as evidence against the existence of any general personality characteristic of suggestibility or conformity.

REFERENCES

1. ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men*. Pittsburgh: Carnegie Press, 1951. Pp. 177–190.
2. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
3. SCHACHTER, S. Deviation, rejection and communication. In L. Festinger *et al.*, *Theory and experiment in social communication*. Ann Arbor: Univer. of Michigan, Research Center for Group Dynamics, Institute for Social Research, 1950. Pp. 51–82.
4. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, **27**, No. 187.
5. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.

Received February 13, 1953.

POLITICAL ATTITUDES AND JUDGMENTS OF OTHER PEOPLE

GUSTAV JAHODA

University College of the Gold Coast[1]

DURING the last few years there have been several studies dealing with judgments of people from pictures. Some of these were concerned with the interrelations of judgments of various attributes (11), or the effects of stereotypes (10). Following recent developments in the theory of perception (4, 8), one sees that interest has increasingly become focused upon the characteristics of the perceiving or judging subjects (5).

This trend has been most conspicuous in connection with ethnic prejudice. Here, the most relevant investigation is that of Razran (9), who asked his subjects to rate girls' photographs on such variables as beauty, intelligence, and character; when these same photographs were later presented with Jewish, Italian, or Irish names, the ratings became modified in accordance with social stereotypes.

Only one other similar work dealing with political attitudes is known to the author. This was an experiment by Hsü (6). He was concerned with judgments of pictures labeled "Communist" and "Non-Communist," respectively (again a minority group), and only three subjects took part.

The present study, carried out in England, was designed to answer the question of whether or not affiliation to one of the two main parties, Conservative or Labour, can exert an influence analogous to that of ethnic prejudice. The aim was to test the hypothesis that, in judging other people, subjects tend to associate favorable personal characteristics with political attitudes of which they themselves approve. More specifically, two effects were predicted: Conservatives and Labourites will tend to make differential judgments of pictures of people they identify as political allies or opponents; and, conversely, they will tend to identify those judged favorably (unfavorably) as belonging to their own (the opposite) party.

[1] The author wishes to acknowledge gratefully the assistance of John Morris and Morris Walker, his former colleagues at the University of Manchester, England, where the research was undertaken.

METHOD

Materials. Photographs were selected from the *Times Guide to the House of Commons*, 1950. This work contains standard-sized photographs of all 625 Members of Parliament (MP's) elected at the 1950 General Election. Most of the pictures were full-face showing the head down to, but excluding, the shoulders.

After eliminating certain categories,[2] batches of 20 photographs equally divided as to party were obtained by means of a table of random numbers. These batches were once more randomly separated into two sets arbitrarily labeled *A* and *B*, each set containing five pictures of Labour (*LAB*) and five of Conservative (*CON*) MP's. The photographs had code numbers so that the experimenters (*E*s) themselves did not know their party.

Criteria. The choice of criteria for the judgments was guided by two considerations: They should involve "values" in the sense of *not* being motivationally neutral, and also should permit the subjects (*S*s) to feel that there was a reasonable possibility of judging a photograph according to the criterion. After several trials, Looks (*LO*), Intelligence (*IQ*), and Social Class (*SC*) were selected. Categories for grading are shown in Table 1.

The exact wording was chosen to insure that none of the resulting distributions would be too skewed. In referring to judgments, the generalized labels "High," "Medium," and "Low" will be used henceforth.

In addition to the above, *S*s were asked to identify party membership for each photograph.

The task. Alternative types of record forms, schematically presented in Table 2, were devised for each criterion separately. The object was both to randomize the sequence of presentation and to separate the judgments of party and criterion.

An example will illustrate the general procedure: *S* is approached with a request to take part in an experiment to study the extent to which individuals agree, and also to determine how accurate they are in judging personality characteristics from photographs. *S* is then given set *A* and asked to sort the pictures into three piles marked, e.g., Upper Class, Middle Class, and Working Class, respectively. When this is completed, *E* records the code numbers in the appropriate spaces on Form I and hands *S* set *B* to sort into two piles according to political party; then set *A* follows, again to be sorted for party; and finally set *B*, to be graded once more according to *SC*.

[2] Liberal MP's, women MP's, and pictures other than full face; MP's in military uniforms or other distinctive dress clearly indicating social status; prominent politicians; MP's from the Lancashire region (where the subjects lived) because they were liable to be recognized.

TABLE 1

CATEGORIES FOR GRADING ON CRITERIA

CRITERION	HIGH	MEDIUM	LOW
LO	good-looking	average	not good-looking
IQ	very high	above average	average and below
SC	upper	middle	working

TABLE 2

ALTERNATIVE MODES OF PRESENTATION

	OBJECT OF JUDGMENT	
SET	Form I	Form II
A	criterion	party
B	party	criterion
A	party	criterion
B	criterion	party

TABLE 3

CONSTITUTION OF THE SAMPLE

	SEX		
PARTY	MALES	FEMALES	BOTH SEXES
Conservatives	49	70	119
Labourites	65	54	119
Others	24	22	46

When *S* has completed the task, he is asked what party he himself would vote for if he had to choose between Conservative and Labour only.

The procedure for Form II was of course the same, apart from the sequence. It will be noted that each *S* was restricted to a single criterion in order to avoid halo effects.

Experimenters and subjects. The field work was carried out by a group of honor students and the author. The *S*s were drawn from the circle of relatives and friends of the experimenters, excluding students except for a few cases where no other contacts were available. The sample procured in this manner, while clearly unrepresentative of the general population, was not confined to any particular social group or class.

The category "Others" in Table 3 is a heterogeneous one, including 21 Liberals, 10 "nonpolitical" subjects, 8 subjects who, on questioning, recognized one or more MP's among the photographs, and 7 persons who refused to divulge political opinion. Except for party identification, these 46 subjects were excluded from the analysis.

RESULTS

Accuracy of party identification. At the outset it is necessary to consider briefly whether recognition of some MP's by *S*s may have affected the results. Before the experiment, trials were undertaken to discover how familiar people were with the faces of MP's. It appeared that individuals were often able to identify their own or neighboring MP's, whose pictures they had seen on boardings or in local newspapers, but hardly anyone recognized other backbenchers. This was to be expected since photographs of backbenchers are rarely published in the national press.

In the actual experiment, less than 10 per cent of the *S*s even guessed that the pictures were of MP's.

Rates of success in identifying party clustered very near to 60 per cent for all groups of *S*s, the only slight deviation being 56.4 per cent for Conservative women. The over-all proportion of correct results was 59.5 per cent, which is significantly better than chance at the .001 level of confidence.

Analysis of judgments. The first question to ask with respect to each of the three criteria was whether the four groups (male and female Conservatives and Labourites) had in fact responded differently to the photographs. If a chi-square analysis proved this to be the case for a particular criterion, successive breakdowns were carried out to discover what accounted for the differences. Within each criterion, the judgments of photographs believed to represent Conservatives (*B/CON*) were treated separately from those believed to represent Labourites (*B/LAB*).

It appears from Table 4 that *LO* and *IQ* were judged differently, but *SC* was not. More detailed breakdowns of the data have shown that the differences exhibited are a function of judgments in which party affiliation was wrongly identified. *LAB* and *CON* judges did

TABLE 4

SIGNIFICANCE OF GROUP DIFFERENCES IN RESPONSES*

	CRITERION		
BELIEVED PARTY	LOOKS	INTELLIGENCE	SOCIAL CLASS
B/LAB	$\chi^2 = 34.06$, $p < .01$	$\chi^2 = 28.67$, $p < .02$	$\chi^2 = 24.96$, ns
B/CON	$\chi^2 = 32.06$, $p < .01$	$\chi^2 = 27.74$, $p < .05$	$\chi^2 = 13.65$, ns

* In all cells $df = 15$

not differ significantly in their judgments of those photographs for which party affiliation was correctly identified.

Distortion of judgments. The following steps were taken in deriving distortion scores. First, a standard judgment of each photograph was obtained by computing the percentages of all *S*s who judged the photograph High, Low, and Medium on a given criterion. In this fashion it was possible to determine what percentage of all judgments of actual *LAB* photographs were High, Low, and Medium on each criterion. Parallel percentages were determined for all actual *CON* photographs. These percentages provided a standard around which distortion scores were computed. The distortion scores were produced by computing the differences (in terms of percentage High and Low judgments) between the standard and the percentage of High and Low judgments given by a segment of the sample (e.g., male Conservatives) to those photographs they believed were representative of affiliation with a given party. Thus in Table 5, male Conservatives are shown to have given 16 per cent fewer High, or more Low, *LO* judgments to photographs they believed to be *LAB* than were given by all *S*s to actual *LAB* photographs.

In the left half of Table 5 the scores of Conservative men and women appear to manifest a consistent bias. This is shown as fewer High judgments of Looks, IQ, and Social Class than the standard for photographs of MP's believed by these *S*s to be of the Labour Party, and more High judgments for photographs of MP's believed to be Conservative.

Labour men exhibit the opposite bias as far as criteria *LO* and *IQ* are concerned, but when it comes to *SC* they deviate in the same direction as the Conservatives. Apart from their moderate amount of distortion, Labour women present no clear-cut trend; regarding *SC*, their responses are uniform with those of all other groups of judges.[3]

Finally, the experimental design provided an answer to the question of whether or not the order of presentation (party first or criterion first) made any difference. In other words, was there a greater tendency for, say, Conservative judges to rate as Low the photographs previously labeled *LAB*, than to pick out those earlier rated, say, "not good-looking" and call them *LAB*? A chi-square test brought out no significant differences, and it may be concluded that in either direction the association was of approximately equal strength.

TABLE 5

DISTRIBUTION OF DISTORTION SCORES

Believed Party	Criterion	Conservative Judges		Labour Judges	
		Males	Females	Males	Females
B/LAB	*LO*	−16.0*	−12.0	+13.0	+3.0
	IQ	−11.0	−13.9	+12.5	−8.5
	SC	−10.0	−21.0	−12.5	−1.0
B/CON	*LO*	+18.5	+10.5	−18.5	−0.5
	IQ	+9.0	+18.0	−6.5	0.0
	SC	+8.5	+13.5	+18.0	+6.5

* This reads, e.g., that male Conservative judges gave 16% less High and/or more Low judgments than the standard. The standard is based on the percentage of judgments from all *S*s that were High or Low for this Looks criterion.

DISCUSSION

The 59.5 per cent success rate in identifying party membership was of the same order of magnitude as those obtained by Allport and Kramer (1) and Lindsey and Rogolsky (7), whose *S*s were required to pick out pictures of Jews. This appears surprising in view of the fact that *LAB* and *CON* MP's as groups, probably differ less from each other than the rank and file supporters of the two parties.

The results of the experiment confirm the general hypothesis stated at the outset. Two specific effects were also postulated: first, overrating of photographs of one's own and underrating those of the opposite party when party identification has come first; second, identifying those previously judged High as belonging to one's own, and those judged Low, to the opposite party. At first sight it appears that both these effects were established, an impression reinforced by the finding that order of presentation (party or criterion first) made no difference.

However, the crucial fact must be emphasized that significant differences occurred only when party was *wrongly* identified. This means that subjects of different parties agreed substantially in their ratings of the photographs, but disagreed on the parties to which they thought the MP's belonged. In other words,

[3] The fact that all *SC* distortions occurred in the same direction accounts for the absence of significant group differences with respect to this criterion.

contrary to appearances, and to the plan of the experiment, judgment on a criterion was always the independent, and party identification the dependent, variable.

In attempting to explain this, a distinction can be drawn between the two operations of judgment on a criterion such as "Looks," and party identification. The former may be called a first-order judgment, being made with reference to directly perceived characteristics; party identification may be termed a second-order judgment, for it is based on a previous first-order one. Thus, in identifying party, the subjects first had to consider the directly observable attributes of the picture and then proceed to an inference or guess concerning party affiliation. Consequently, whatever the order of presentation, implicit first-order judgments, albeit not necessarily on a specified criterion, must have come first. If these considerations are justified, it follows that the experimental design was unsuitable for testing the first of the effects above; it could only have been done by providing subjects with ready-made party labels for the photographs.

What did emerge conclusively was that Labour men gave a disproportionate share of the good-looking and intelligent-appearing types to the Labour Party, while Conservatives, and to some extent Labour women, assigned an excessively large number to the Conservative Party. The distortion scores, of course, have to be interpreted accordingly.

So far, no reference has been made to the discrepancy between the response patterns relating to *LO* and *IQ* on the one hand, and to *SC* on the other. This may be partly accounted for, perhaps, by the fact that "favorable" and "unfavorable" are somewhat ambiguous terms when applied to social class levels in England, for valuation depends to some extent on the individual's position within the class hierarchy and his political attitudes. For instance, Conservative *S*s tended to characterize the working class by such epithets as "coarse," "shabby," "dull"; Labourites again referred to upper-class people as "snobbish," "smug," "hard-faced."

It is unlikely, however, that such differential class valuations are sufficient to explain the observed discrepancy; a more general hypothesis may be put forward tentatively to account for the data as a whole: in their relation to the outer world, people expect certain attributes to occur together. Their expectations may be based either on motivational factors, or upon reality, or both. Whenever people encounter new situations or new people, their judgments will be affected by these expectations which they carry with them into the situation. With respect to *LO* and *IQ*, Conservative and Labour subjects produced different patterns of distortion because their expectations differed; with respect to *SC*, the distortion patterns were similar to one another because reality had created in both types of subject the expectation that Conservatives are likely to have high, and Labourites low, class status.

While no direct data were collected on the various subjects' expectations, their comments in the course of the experiment tend to support the above hypothesis. Statements such as the following were frequent with Conservative subjects: "The ones with breeding in their features are Conservatives"; "Socialists are rough-looking types and lack polish"; "The ones with a vacant expression are Labour." And on the other side: "The fat and stupid-looking ones are Conservatives"; "Labour people have a frank and open appearance." Similar descriptions of working-class people and Labourites, or upper-class people and Conservatives, often implied a bracketing of the two pairs of categories.

There remains the problem of the deviant responses of Labour women. The relevant literature (2, 3, 12) contains a good deal of evidence to show that Labour and/or working class women diverge considerably from Labour men in their social attitudes. The divergence is mostly in the Conservative direction, and one is led to conclude that a willingness to vote Labour rests on a somewhat different basis in women from that usual in men.

Finally, when assessing the extent of the distortions obtained in accordance with the initial hypothesis, it should be noted that the results might have been more striking in the absence of two negative factors: first, some of the more sophisticated subjects were aware of the possibility of bias inherent in the judging situation and attempted to guard against it; second, the subjects were not specially selected for the intensity of their attitudes, which ranged all the way from enthusiasm to rather half-hearted support.

SUMMARY

1. The hypothesis was set up that in judging other people, subjects tend to associate favorable personal characteristics with political attitudes of which they themselves approve.

2. In order to test the hypothesis, a set of photographs of British Members of Parliament was graded according to three criteria by 119 Conservative and 119 Labour subjects. A further task was the party identification of the men pictured.

3. About 60 per cent of the photographs were correctly identified as to party, a result considerably better than chance.

4. There were significant differences between Conservative and Labour subjects with respect to criteria Looks and Intelligence, but not to Social Class. These differences occurred only when photographs were wrongly identified in terms of party.

5. With reference to Looks and Intelligence, distortions were found in the expected direction, favorable to political allies and unfavorable to opponents.

6. The responses of men and women were similar with Conservatives, but divergent in the case of Labour supporters.

REFERENCES

1. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9–39.
2. BENNEY, M., & GEISS, P. Social class and politics in Greenwich. *Brit. J. Sociol.*, 1950, **1**, 310–327.
3. BIRCH, A. H., & CAMPBELL, P. Voting behaviour in a Lancashire Constituency. *Brit. J. Sociol.* 1950, **1**, 197–208.
4. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. Blake & G. Ramsey (Eds.), *Perception, an approach to personality*. New York: Ronald, 1951. Pp. 121–147.
5. FENSTERHEIM, H., & TRESSELT, M. E. The influence of value systems on the perception of people. *J. abnorm. soc. Psychol.*, 1953, **48**, 93–98.
6. HSÜ, E. H. An experimental study of rationalization. *J. abnorm. soc. Psychol.*, 1949, **44**, 277–278.
7. LINDSEY, G., & ROGOLSKY, S. Prejudice and identification of minority group membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 37–53.
8. POSTMAN, L. Toward a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951. Pp. 242–272.
9. RAZRAN, G. Ethnic dislikes and stereotypes. A laboratory study. *J. abnorm. soc. Psychol.*, 1950, **45**, 7–27.
10. SECORD, P. F., BEVAN, W., JR., & DUKES, W. F. Occupational and physiognomic stereotypes in the perception of photographs. *J. soc. Psychol.*, 1953, **37**, 261–270.
11. WILKIE, J. S. An experiment on reading character from portraits. *J. gen. Psychol.*, 1949, **40**, 11–45.
12. ZWEIG, F. *Women's life and labour*. London: Victor Gollancz, 1952.

Received March 28, 1953.

REACTION TIME AS A FUNCTION OF MANIFEST ANXIETY AND STIMULUS INTENSITY[1]

CHARLES WENAR

University of Illinois, College of Medicine

THERE is at present considerable agreement between clinical theory and behavior theorists' formulations of the nature of anxiety. Fenichel (2) states: "All anxiety is a fear of experiencing a traumatic state (p. 133) . . . . It is the ego's judgment, 'danger ahead' (p. 213) . . . . The ego's judgment of impending danger brings the organism into a state similar to that of the trauma, but of lower intensity. The 'tamed' anxiety . . . is used to indicate the necessity for starting defensive action" (p. 132). A similar idea is elaborated by Mowrer (8), and by Miller and Dollard (7), who regard anxiety as a conditioned pain response which has drive value. Thus, stimuli present at the time of a traumatic event tend, on being repeated, to arouse certain components of the pain response, even though the painful event itself does not occur. Man, being a highly articulate organism, can report being "afraid" or "anxious" at such times. When such anxiety is prolonged over a considerable period of time, even though no objective dangers can be identified, the clinical term "anxiety state" or "anxiety neurosis" is used. Most descriptions include two types of response from which the anxiety state is inferred: first, there are complaints of generalized fear of varying intensities, depending on the severity of the case; and second, there are concomitant physiological changes such as rapid heartbeat, increased sweating of skin surfaces, etc.

For the present study, the important inference from both clinical theory and behavior theory is that chronically anxious individuals operate at a higher drive level than nonanxious ones. There is also considerable experimental evidence, involving a number of different response measures, which points to the validity of attributing motivational properties to anxiety. Taylor (16) found more rapid eyelid conditioning in anxious subjects than in nonanxious subjects, and Rosenbaum (11) found greater responsiveness to generalized stimuli in an anxious than in a nonanxious group. Both studies interpreted the results as indicating a higher drive level in the anxious group on the basis of Hull's theory (3) that an increase in drive facilitates the evocation of a response. Johanson's findings (5) that reaction time decreased in subjects made anxious by shocking them for reacting slowly can be interpreted in terms of increase in motivation.

PROBLEM

Since there is experimental evidence that not only anxiety but also stimulus intensity may be considered a motivational factor,[2] the problem arises concerning the effect on behavior of varying both the anxiety level of individuals and the intensity of the stimulus. Hull's theoretical approach states this problem precisely enough to lead to predictions which can be experimentally verified. According to him, if the learning factor ($\hat{s}H\hat{k}$) is held constant, then the higher the drive level, the greater the magnitude of the response tendency. From this, one would predict that anxious individuals would react more strongly than nonanxious individuals to discrete stimuli in simple situations. However, he also postulates that the total drive is composed of irrelevant ($\dot{D}$) and relevant (D) drives which summate according to the formula (100) $\dot{D} + D/\dot{D} + 100$. Now, in an experimental situation in which anxious and nonanxious individuals respond to discrete stimuli, anxiety can be considered as contributing to the irrelevant drive, and the intensity of the experimental stimulus can be considered as contributing to the relevant drive. On these assumptions the application

[1] This study was part of a doctoral dissertation completed in 1951 in the department of psychology at the State University of Iowa, under the direction of Dr. I. E. Farber. The author is also indebted to Dr. J. S. Brown and Mr. Ted Hunter for their help in designing and constructing the apparatus.

[2] Passey (9) and Spence and Taylor (14) found more rapid eyelid conditioning with a strong puff of air than with a weak puff; Woodworth (17) cites studies indicating that reaction time decreases as the intensity of stimulation increases. Also, see Spence (13) for a more explicit statement of the way in which intensity of stimulation is now treated in the framework of S-R theory.

of Hull's formula would result in the expectation that the total magnitude of response of anxious individuals to a stimulus should be greater than that of nonanxious individuals, but differences between the two groups should decrease with increasing intensity of stimulation.

Experimental evidence on this problem, however, is far from clear. Some results are in line with the implications of the above-mentioned theory. Malmo (6), using as a response measure the number of "bursts" of motor activity defined in terms of finger movements on a button the subject could press if the stimulus became too painful, found that anxious subjects reacted more strongly than nonanxious subjects to a mild pain stimulus, and that they continued to do so as the intensity of the stimulus was increased. However, the rate of increase was more rapid in the nonanxious group. Solomon and Fentress (12) found that anxious individuals had low resistance levels initially and had small absolute changes in GSR to different sensory stimuli such as a buzzer and a light; nonanxious individuals had high resistance levels and large absolute changes. Darrow and Solomon (1), using a similar procedure with psychotic patients, found similar results in regard to absolute change in GSR.

Other experimental findings are at variance with predictions from the present application of Hull's formula. Spence and Taylor (15) using a weak and a strong puff of air in conditioning the eyelid response found that the speed of conditioning of anxious and nonanxious individuals was not significantly affected by increasing the intensity of the noxious stimulus. Rosenbaum (11) found that the difference in stimulus generalization between anxious and nonanxious individuals was significant only when a strong shock was administered for a given number of incorrect responses; when a mild shock or a buzzer was used, there was no significant difference between the two groups. This led him to speculate that the drive properties of clinical anxiety may not be activated under conditions of little or no threat.

Hypotheses

1. On the assumption that anxiety can be considered as contributing to the irrelevant drive ($\dot{D}$), and the drive induced by the intensity of the experimental stimulus can be considered as contributing to the relevant drive (D), Hull's formula for drive combination would imply that the total drive level of anxious groups would be higher than that of nonanxious groups; however, increase in stimulus intensity should result in a reduction of the difference between the total drive level of the anxious and nonanxious groups.

2. If Rosenbaum's speculation that the drive properties associated with clinical anxiety may not be aroused under experimental conditions of little or no threat is correct, there should be no difference in the drive level of anxious and nonanxious individuals when the stimulus is relatively mild and innocuous.

METHOD

Design. The above hypotheses were tested by measuring the reaction time of anxious and nonanxious individuals to three different stimuli—a buzzer, a weak shock, and a strong shock. With other factors held constant, it was assumed that speed of reaction (1/R.T.) would be a positive function of the variations in motivation associated with variations in level of manifest anxiety and stimulus intensity. In the initial or "training" series, the stimuli were presented at regular 12-sec. intervals and, on the above assumption, increased speed of reaction during training was used as one criterion of increased drive. Following training, the subject's reaction time to stimuli presented at intervals of shorter and longer duration than the training interval was determined in order to obtain a temporal gradient of response strength. On the assumption that such a temporal gradient of response strength would represent a special instance of stimulus generalization,[2] it was supposed that an increase in drive, in terms of anxiety and/or stimulus intensity, would raise the height and steepen the slope of the obtained gradient. This, then, was the second criterion of increased drive.

Subjects. Subjects (*Ss*) were members of the under-

[2] The rationale here derives from Hull who postulates that the generalization of habit strength from a conditioned stimulus to other stimuli along a given sensory dimension is a negative growth function of the distance, in sensory units, between these stimuli and the conditioned stimulus. Since he also postulates that drive multiplies all existing habit systems, it follows that when drive is increased, the response tendency to the conditioned stimulus and to stimuli differing from it along a given dimension should be strengthened. This produces both an increase in the height and a steepening of the slope of the generalization gradient. Rosenbaum (10) pointed out that the temporal gradient of response strength can be regarded as a special instance of empirical stimulus generalization in which habit strength generalizes to internal cue patterns rather than to external sensory stimuli.

graduate course in introductory psychology at the State University of Iowa. Anxious and nonanxious individuals were selected by means of a revision of the Anxiety Scale devised by Taylor (16). This scale consists of 50 items from the Minnesota Multiphasic test which were judged by trained clinicians to be the clearest indices of symptomatic anxiety. Forty-two *Ss* were taken from those obtaining scores in the upper 20 per cent of the distribution, and an equal number from the lower 20 per cent.

Apparatus. For individuals in the shock groups the output voltage of a Variac, which was connected to the 115-volt, 60-cycle supply line, was fed to the *Ss* through a pair of isolating transformers. Dry electrodes consisting of smoothed-surface dimes were used. The voltages applied to the *Ss* from these transformers were read from a 2000 ohm-per-volt AC voltmeter. For individuals in the buzzer group, a buzzer of medium intensity was placed behind a cardboard shield approximately 8 in. from the reaction-time switch. In the case of both shock and buzzer, the stimulus was of brief duration.

Two Hunter-Brown Electronic Timers (4) were used to regulate the interval between presentations of the stimulus. They were connected so that the dial could be set for the standard 12-sec. interval and reset for the desired interpolated intervals.

In order to measure reaction time, a Standard Electric clock was started the instant the stimulus was presented and was stopped when the subject depressed a lever on a microswitch.

All apparatus except the reaction-time switch were screened from the subject's view.

Procedure. Both anxious and nonanxious groups were subdivided into three groups of 14 *Ss* according to the type of stimulus used as a signal for reaction—a mild shock, a strong shock, or a buzzer. Thus there were six groups comprising a factorial design with two degrees of anxiety and three intensities of stimulation. Anxious and nonanxious individuals were assigned to the groups at random, and all six conditions were run concomitantly.

At the beginning of the experiment *S* was seated at a table with the reaction-time key on his right and (in the case of the shock groups) the electrodes attached slightly above the wrist of his right arm. In order to determine the voltage necessary to produce a "strong" and a "weak" shock, each *S* was given a series of shocks, beginning at 10 volts and increasing to 50 volts, which he was instructed to rate as mild, medium, or strong. The mean value of shock for both Strong Shock Groups was 26.6 volts; the mean value for both Weak Shock Groups was 17.9 volts. There was no tendency for one anxiety group to be systematically lower or higher than the other in its judgments of mild and intense voltages. For all groups, the strength of shock was kept constant throughout the experiments.

All *Ss* were given three or four practice trials in pressing the reaction-time key in order to accustom them to the movement. In the case of the Buzzer Groups, the buzzer was presented three times before the experiment proper began so that the *S* might readily identify the signal. Instructions were worded so as to emphasize the speed of the response, and *S* was told his reaction time after each presentation of the stimulus in order to maintain his motivation.

In the initial phase of the experiment, the stimulus was delivered at regular 12-sec. intervals. There were 17 such presentations, which constituted the training or temporal conditioning series. Then, on trials 18, 21, 27, 35, 41, 48, and 55 the signal was presented at varying intervals. These intervals were 3, 6, 9, 12, 15, 18, and 21 seconds, and their sequence of presentation was varied in terms of a latin-square design. These intervals constituted the test trials for determining the temporal gradient of response strength. Between the test trials, the 12-sec. interval was administered in order to reestablish the original temporal conditioning.

The response measure for reaction time was the time, in hundredths of a second, between the administration of the stimulus and the depression of the key.

RESULTS

Before analyzing the data, a reciprocal transformation was performed on all reaction-time scores in order to eliminate the typical skewness of the distribution of latency measures. Because of this the response measure will be referred to as speed of reaction rather than reaction time.

Speed of reaction during training. The first 17 trials were divided into four blocks of trials, and the mean speed of reaction per trial within each block was plotted. Since the resulting curves appeared to be approximately parallel, the mean of the speed-of-reaction score for the 17 trials was used as a representative measure of the speed of reaction of the various subgroups during training. These means of speed-of-reaction scores for each subgroup are presented in Table 1. An analysis of variance indicated that both increase in anxiety and increase in stimulus intensity were effective in increasing the speed of reaction during training. The former was significant at the 5% level of confidence, the latter at the .1% level of confidence. The interaction between anxiety and stimulus intensity was not significant.

Height of the temporal gradient of response

TABLE 1

MEAN OF THE SPEED-OF-REACTION SCORES FOR TRIALS 1—17 ($N = 14$)

GROUPS	STRONG SHOCK	WEAK SHOCK	BUZZER
Anxious	7.62	5.86	7.25
Nonanxious	7.16	5.36	6.07

TABLE 2

MEAN OF THE SPEED-OF-REACTION SCORES AT THE 3-, 6-, AND 9-SEC. INTERVALS

GROUPS	STRONG SHOCK	WEAK SHOCK	BUZZER
Anxious	5.10	4.15	4.29
Nonanxious	5.01	3.49	3.75

strength. The temporal gradient of response strength was obtained by plotting the mean speed of reaction for each group at each of the test intervals. Since the curves for the anterior portion of the gradient (intervals of 3, 6, and 9 sec.) appeared to be quite different from those for the posterior portion (intervals of 15, 18, and 21 sec.), the two segments were analyzed separately.

As a measure of the height of the anterior portion of the gradient, the mean speed-of-reaction scores at the 3-, 6-, and 9-sec. intervals were summated for each group. The mean of these scores for each group is presented in Table 2. An analysis of variance indicated that an increase in anxiety significantly increased the height of the anterior portion of the temporal gradient of response strength (at the 5% level of confidence), and that an increase in stimulus intensity also increased the height significantly (at the .1% level of confidence). The interaction between anxiety and stimulus intensity again was not significant.

As a measure of the height of the posterior portion of the gradient, the mean speed-of-reaction scores at the 15-, 18-, and 21-sec. intervals were summated for each group. The mean of these scores for each group is presented in Table 3. An analysis of variance indicated that both increase in anxiety and increase in stimulus intensity were effective in increasing the height of the posterior portion of the temporal gradient of response strength. The former was significant at the .1% level of confidence, and the latter was significant at the 1% level of confidence. The interaction between anxiety and stimulus intensity was not significant.

TABLE 3

MEAN OF THE SPEED-OF-REACTION SCORES AT THE 15-, 18-, AND 21-SEC. INTERVALS

GROUPS	STRONG SHOCK	WEAK SHOCK	BUZZER
Anxious	8.49	7.11	7.71
Nonanxicus	7.38	5.73	6.33

TABLE 4

MEAN OF DIFFERENCES OBTAINED BY SUBTRACTING MEAN SPEED-OF-REACTION SCORES AT 3 SEC. FROM MEAN SPEED-OF-REACTION SCORES AT 12 SEC.

GROUPS	STRONG SHOCK	WEAK SHOCK	BUZZER
Anxious	3.75	5.99	5.14
Nonanxious	4.37	3.95	4.38

Slope of the temporal gradient of response strength. As a measure of the steepness of slope of the gradient for intervals shorter than the training interval, the speed of reaction at the 3-sec. interval was subtracted from the speed of reaction at the 12-sec. interval. The mean of these differences for each group is presented in Table 4. An analysis of variance indicated that neither increase in anxiety nor increase in stimulus intensity was effective in steepening the slope of the anterior portion of the temporal gradient of response strength. The interaction between anxiety and stimulus intensity likewise was not significant.

An examination of the posterior portion of the temporal gradient of response strength revealed that, for the Anxious Strong Shock Group, the speed of reaction showed an increase from the 12-sec. to the 21-sec. interval, whereas, in all other groups, the speed of reaction decreased. Because of the theoretical implications of such a finding, if reliable, an analysis of trend of the six curves was done in order to provide a more sensitive measure of possible differences in slope between groups. The results indicated that neither increase in anxiety nor increase in stimulus intensity was effective in steepening the slope of the posterior portion of the temporal gradient of response strength.

Since the steepness of the slope of the temporal gradient of response strength was the only instance of insignificant results, the gradients were examined for a possible explanation. An examination revealed that in the case of the Anxious Strong Shock Group the speed of reaction, instead of increasing appreciably at the 12-sec. interval, was only slightly faster than at the 9-sec. interval. For

every other group there was a considerable rise in the gradient between these two points. In addition, it was not the over-all shape of the gradient of the Anxious Strong Shock Group which deviated from that of the other groups, but only the reaction at the 12-sec. interval. It would seem a reasonable inference that this point represented a chance deviation rather than a stable characteristic of the group. On logical grounds there is no apparent reason why the relative positions of the groups in terms of speed of reaction at this 12-sec. interval should have differed from their positions in the training trials which were also 12 sec. apart. If this score of the Anxious Strong Shock Group was a chance deviation, the results of the above analyses of differences in steepness of slope of the temporal gradient of response strength are unreliable, and conclusions from them should be drawn only with great caution.

DISCUSSION

Since speed of reaction was used as a measure of drive in the present study, the results provided additional evidence that both manifest anxiety and stimulus intensity may be regarded as having motivational effects. The results obtained indicated that either an increase in anxiety level or an increase in the intensity of stimulation may produce a significant increase in drive.

In regard to the problem of the effect of different stimulus intensities on the drive of anxious and nonanxious individuals, there was no evidence for the predicted decrease in the difference in total drive levels of the two groups as the stimulus intensity was increased. However, there are two important considerations. First, the use of both a buzzer and a shock as signals involved differences not only in the intensity of the stimuli, but in the modality affected. Since there is evidence indicating that reaction time is slower in the case of shock than in the case of auditory stimuli, the issue of how much the increase in speed of reaction in the present study was due to increased intensity of the stimulus and how much it was due to differences in modalities is confounded. Second, examination of the scores of the two intensities of shock alone for the two criteria of increased speed of reaction during training and raising the temporal gradient reveals that the decrease in difference between anxious and nonanxious groups was in the predicted direction. It is possible that the introduction of a stimulus affecting another modality might have obscured this relationship to some extent.

The theoretical explanation of the failure to get the predicted decrease in difference in total drive level between anxious and nonanxious groups with increased stimulus intensity which is simplest and most in keeping with the observed behavior of the subjects may be that, in spite of the fact that the anxious groups consistently reacted faster than the nonanxious ones, the irrelevant anxiety drive was fairly high for both groups. For example, if one thinks of the anxiety drive as ranging from unconcern (no anxiety) to panic (extreme anxiety), it is possible to have one group significantly less anxious than another but, at the same time, to have both groups fairly high on this range. Thus, it is quite possible to have one group less anxious than another but still tense enough to prevent a significant decrease in difference between the two groups with increased intensity of stimulation.[4] Since 64 of the 84 *S*s had been in an experiment in which they were shocked, and since most of them expressed concern that they would be shocked again, the entire population might be described as "shock sensitive."

The present results also indicate that, contrary to Rosenbaum's speculation that the drive properties of anxiety may not be aroused under conditions of little or no threat, anxious individuals had a higher drive level than nonanxious ones even when the stimulus was not very intense. Such findings fit nicely with the clinical concept of anxiety. What the anxious individual often fears most is a new or unfamiliar situation; that is, he has a generalized fear of "what might happen." No matter how innocuous the task, therefore, the experimental situation itself is often sufficiently threatening to arouse a significantly greater amount of manifest anxiety in him than in nonanxious individuals.

[4] If one thinks of the drive value of anxiety and of stimulus intensity as ranging from 0 to 100, Hull's formula for combination of drives, $(100)\ \dot{D} + D/\dot{D} + 100$ can be used to construct a table of hypothetical values of total drive to illustrate the above point.

SUMMARY

The present study was designed to investigate the relative increase in drive of anxious and nonanxious individuals as a function of different intensities of stimulation. On the assumption that, in a reaction-time study, anxiety can be considered as contributing to the irrelevant drive and stimulus intensity can be considered as contributing to the relevant drive, Hull's formula for drive combination was used to predict that the total drive level of anxious individuals would be greater than that of nonanxious individuals, but that an increase in stimulus intensity would result in a reduction in the difference in the total drive level between the two groups.

Anxious *Ss* were defined as individuals whose scores were in the upper 20 per cent of the Taylor anxiety scale, and nonanxious *Ss* as individuals whose scores were in the lower 20 per cent of the scale. There were 42 *Ss* in each group. Stimulus intensity was varied by using a strong shock, a weak shock, and a buzzer as signals for pressing a key in a simple reaction-time procedure. Following training, reaction time to stimuli presented at intervals of shorter and longer duration than the 12-sec. training interval was determined in order to obtain a temporal gradient of response strength. Increased speed of reaction during training and increase in the height and a steepening of the slope of the temporal gradient of response strength were the three measures of increase in drive.

The results indicated that both an increase in manifest anxiety and an increase in stimulus intensity were effective in increasing the speed of reaction during training, and in raising the height of the temporal gradient of response strength; neither factor was effective in steepening the slope of the gradient. In none of the measures used was there a significant change in the difference in reaction time between anxious and nonanxious groups as a function of differences in stimulus intensity. These results were interpreted as indicating that both increase in manifest anxiety and increase in stimulus intensity may be regarded as having motivational effects. It was speculated that the failure to find a significant decrease in the difference in speed of reaction between anxious and nonanxious individuals with increased stimulus intensity might be attributable to the fact that both groups were responding at relatively high anxiety levels and/or to the fact that the stimulus not only varied in intensity but also affected different modalities.

REFERENCES

1. DARROW, C. W., & SOLOMON, A. P. Galvanic skin reflex and blood pressure in psychotic states: reaction to sensory, indifferent ideational and critical ideational stimuli. *Arch. Neurol. Psychiat., Chicago*, 1934, **32**, 273–299.
2. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
3. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
4. HUNTER, T. A., & BROWN, J. S. A decade-type electronic interval timer. *Amer. J. Psychol.*, 1949, **62**, 570–575.
5. JOHANSON, A. M. Influence of incentive and punishment upon reaction time. *Arch. Psychol.*, 1922, **8**, 54–93.
6. MALMO, R. B. Experimental studies of mental patients under stress. In M. L. Reymert (Ed.), *Feeling and emotion*. New York: McGraw-Hill, 1950.
7. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
8. MOWRER, O. H. Stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
9. PASSEY, G. E. The influence of the intensity of unconditioned stimulus on the acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38**, 420–428.
10. ROSENBAUM, G. An experimental study of a temporal gradient of response strength in instrumental conditioning. Unpublished master's thesis, Univer. of Iowa, 1947.
11. ROSENBAUM, G. Stimulus generalization as a function of clinical and experimentally induced anxiety. Unpublished doctor's dissertation, Univer. of Iowa, 1950.
12. SOLOMON, A. P., & FENTRESS, T. L. Galvanic skin reflex and blood pressure reactions in psychoneurosis. *J. nerv. ment. Dis.*, 1934, **80**, 163–182.
13. SPENCE, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.), *Comparative psychology*. New York: Prentice-Hall, 1951.
14. SPENCE, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.
15. SPENCE, K. W., & TAYLOR, J. Anxiety and the strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
16. TAYLOR, J. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
17. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

Received March 7, 1953.

SOME EFFECTS OF COOPERATION AND COMPETITION UPON SMALL GROUP BEHAVIOR[1]

MARTIN M. GROSSACK

Philander Smith College, Little Rock, Arkansas

COOPERATIVE and competitive social situations are often said to have differential effects on group behavior. This paper reports an experiment testing specific hypotheses that concern some of these differences. The theoretical rationale is based on the following three assumptions:

1. A cooperative situation is one of "task direction" (1), while a competitive situation is not. A situation of task direction is defined as one in which an individual goal is reached by the achieving of a group's task.

According to Deutsch (2), the basic difference between cooperation and competition is in the way an individual achieves his goal in both situations. Cooperation was defined by him as a situation in which no individual reaches his goal unless all other individuals also enter their goal regions. Competitive situations were defined as those in which no individual reaches his goal unless all other individuals are unable to enter their goal regions.

2. Cooperation may be viewed as an example of cohesiveness due to task direction. Back (1) demonstrated that a relatively high amount of cohesiveness ("total field of forces which act on members to remain in the group") existed in situations where the group mediated individual goals.

3. Conditions of task direction (e.g., cooperation) are likely to produce "instrumental communications," while competitive situations will tend not to have such an effect. Festinger (4) described instrumental communications as those "in which the reduction of the force to communicate depends upon the effect of the communication on the recipient."

Competition, as defined here, requires that an individual's success lead to the failure of other group members. Instrumental communications (influencing another member to accept an accurate solution to a problem, exerting pressures toward uniformity or communicating relevant information) would serve to minimize a competitive individual's own chance of reaching his goal.

HYPOTHESES

1. Indications of cohesiveness will be more frequent among cooperative than among competitive individuals.

a. Cooperative individuals will attempt to influence others more than competitive individuals will.

This prediction is based on the second assumption. It has empirical substantiation in findings reported by Back (1) and Festinger, Schachter, and Back (5) that cohesiveness is a determinant of social influence, and has the following theoretical rationale:

> The magnitude of change the group can induce (its internal power) will be equal to or less than the magnitude of the resultant force on the member to remain in the group (its cohesiveness) (5, p. 166).

b. Cooperative individuals will send and accept more pressures toward uniformity than will competitive individuals.

This hypothesis is derivable from Festinger's theoretical treatment of communication (4). Results of Festinger and Thibaut (7) show more acceptance of pressures toward uniformity when greater pressures are induced on individual members. In a cooperative situation we can assume such pressures to exist because of the need to locomote as a group.

c. Cooperative individuals will send more relevant communications than will competitive individuals.

Communications of relevant information may be considered instrumental for cooperative group locomotion, but not for individuals in a competitive situation. Interpersonal communication of an instrumental type may be considered the means by which cooperative individuals facilitate one another's goal achievement.

[1] This article is based on a dissertation submitted to the graduate school of Boston University in 1952. A detailed review of the literature may be found in that volume (8). The writer acknowledges the constructive help of Drs. Robert Chin, Nathan Maccoby, and John Thibaut throughout this research.

2. Cooperative individuals will send and receive more communications, more instrumental communications, and fewer consummatory communications than will competitive individuals.

Communications of opinion and information are considered instrumental, while those of tension and antagonism are coordinated to consummatory communications (those in which "the reduction of the force to communicate occurs as a result of the expression and does not depend upon the effect . . . on the recipient" [4, p. 281]).

a. Cooperative individuals are more likely to take group functions (actions that facilitate group solidarity or smooth functioning) and less likely to take individual functions (actions not directed toward group goal achievement, e.g., aggression, blocking, etc.) than are competitive individuals.

This is a replication of Deutsch's hypothesis (2).

METHOD

Eighteen female student groups of five individuals each were created for experimental study. Seats in the experimental room were so arranged that subjects (*S*s), upon entering and receiving places, had to face a wall, and were unable to see one another.

The participating *S*s were asked neither to speak nor turn around and to raise their hands if there were any questions. After five *S*s were seated and assigned letters from A to E, the experimental group meeting was ready to begin.

Creation of Independent Variable

Each *S* received a sheet containing a different set of instructions except for the fifth person, who was given one of the other four randomly. The types of instructions distributed were:[2]

1. ***High motivation, cooperative.*** The *S*s were told that the *group* would be rated as a *group* on its solutions to a human relations problem. The best *group* would be rewarded.

2. ***High motivation, competitive.*** Here *S* was told that she would be rated *individually* and that the *individual* with the best solution would be rewarded.

3. ***Low motivation, cooperative.*** The *S* was told that *E* was just studying how *groups* solve problems, that *S*s will be rated as a *group* but that no one will ever know about the group's rating.

4. ***Low motivation, competitive.*** The *S* would be rated as an *individual*. The *E* was described as interested in how *individuals* solve problems, but not interested in whether any specific person does well or poorly.

(The motivational aspects of the instructions were quite unsatisfactory in producing differences in response. An analysis of the data in terms of motivational effects was not completed.)

The *S*s were permitted to communicate by writing notes, if they wanted to, during three periods of four minutes each.

Group Task

The problem selected for use in this experiment was a short case study of a delinquent boy, Johnny Rocco (12). Individuals receiving the case study were asked to consider the problem from the point of view of a person responsible for the boy's future treatment. Schachter's love-punishment continuum, modified by the writer, was also distributed. The *S*s were asked three times during the experiment to indicate which of seven alternatives was the preferred treatment for the boy.

Control of Group Process

It was believed methodologically valuable to develop a technique that would control the operation of personality variables in the interaction situation. By systematically substituting a standard set of stimulus messages, designed by *E*, for those actually sent by the *S*s, control of this aspect of interaction became possible. Also, this prevented individuals from learning that each received different instructions. Some implications of this technique are discussed elsewhere (9).

Each *S* received a set of identical notes, supposedly written by the other four people in the group. The fifth individual was necessary so that each *S* would receive the same amount of notes (i.e., four). The standard stimulus messages had the following thematic content:

From A: Shows high motivation to work on the group's problem, wants to work together with others in the group and to reach a common solution.

From B: Shows high motivation and desire to work on the problem alone.

From C: Shows low motivation, is in field to the extent that communication mentions the problem and the "sender's" attitude toward it.

From D: Shows low motivation and tends to be out of field, horseplays.[3]

The experimenter collected messages written by the subjects; meanwhile, *S*s indicated preferences on the love-punishment continuum and reread the experimental instructions. Then *S*s were permitted to open the stimulus notes received. During the last writing period, however, *S*s *first* read the planted notes and *then* indicated their preferences on the love-punishment scale. A postsession questionnaire consisting of eighteen "agree-disagree" items, three sociometric, and one concerning the instructions was then distributed.

[2] The full text of the instructions has been deposited with the American Documentation Institute. Order Document No. 4174 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 x 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[3] The notes are more fully described in (10).

Summary of Experimental Procedure

For a convenient view of the experimental procedure, the steps in each group may be summarized as follows:

1. Five *S*s, facing the wall, were asked neither to communicate with nor turn towards one another.
2. Introductory remarks are made to put *S*s at ease.
3. Receipt of experimental written instructions (independent variable). Some *S*s receive cooperative, others competitive instructions.
4. Receipt of human relations problem and scale of suggested solutions to it. Instructions indicate that communications are restricted to written messages during three note-writing (N1, N2, N3) periods. The *S*s are to communicate at their own discretion.
5. First note-writing period (N1). The *S*s write notes which are collected, then indicate preference on scale, are reminded neither to turn nor communicate, reread instructions, and receive set one of standard stimulus notes.
6. The *S*s read received planted notes (set one), second note-writing period (N2) starts, *S*s write notes which are collected, then indicate preference on scale, are reminded neither to turn nor communicate, reread instructions, and receive set two of planted messages.
7. The *S*s read received planted messages (set two), third note-writing period starts (N3); *S*s write notes which are collected, receive set three of planted notes, read notes, and indicate preferences on love-punishment scale.
8. Distribution of postsession questionnaire which *S*s answer.
9. Participants are told about the experiment and all questions answered when possible. They are thanked for their cooperation and asked to keep the details of the experiment confidential.

RESULTS AND INTERPRETATIONS

The results of this experiment will be discussed in four sections: (*a*) the experimental instructions, (*b*) cooperation as a determinant of cohesiveness, (*c*) cooperation as a determinant of instrumental communication, and (*d*) conflicts in expectations.

The Experimental Instructions

The success of the experiment as a whole was contingent upon the efficacy of the experimental instructions. A measure of the impressions given by the instructions was available in responses to the last questionnaire item: "At the time I received the instructions, they gave me the following impression (check one):

1. The problem isn't important for me, I will be rated as my group is rated.
2. The problem isn't important for me, I will be rated as an individual.
3. The problem is important for me, I will be rated as my group is rated.
4. The problem is important for me, I will be rated as an individual." (Table 1)

TABLE 1

DIFFERENCE BETWEEN RECIPIENTS OF COOPERATIVE AND COMPETITIVE INSTRUCTIONS IN FREQUENCY OF RESPONSE TO QUESTIONNAIRE ITEM*

INSTRUCTION RECEIVED	RESPONSE	
	GROUP	INDIVIDUAL
Cooperative	37	8
Competitive	10	34

* Chi square = 28.9, $p = .001$.

The results in Table 1 support the point of view that the cooperative and competitive instructions had consequences for individual perceptions in the expected direction. Of the 45 *S*s receiving cooperative instructions, 37 reported the feeling that they would be rated as a group. Of 44 *S*s receiving competitive instructions, 34 reported the feeling that they would be rated as individuals (one *S* did not answer the item in a scorable manner). Such differences could be expected to occur by chance less than .001.

More evidence for the successful creation of cooperation and competition comes from responses to item 8 of the postsession questionnaire as reported in Table 2. The statistically significant difference in responses to that item ("My success in the group was pretty dependent on the other members also succeeding") indicates that the instructions were effective as far as cooperation and competition are concerned.

The causal perceptions for experiences of success and failure in the group meeting reflected the experimentally created social situation. Cooperative *S*s agreed with item 15, "If I do not get a rating of one, it will be due mainly to the other people in the group" significantly more than did recipients of competitive instructions. Similarly, item 14, "If I get a rating of one, it will be due mainly to my own efforts" was agreed to by competitive *S*s significantly more often than by cooperative *S*s (Table 2).

Cooperation as a Determinant of Cohesiveness

It was predicted that more cohesiveness would be shown by cooperative than by competitive individuals. This was expected on the

TABLE 2

AVERAGE DIFFERENCES BETWEEN COOPERATIVE AND COMPETITIVE SUBJECTS IN RESPONSE TO QUESTIONNAIRE ITEMS

ITEM	COOP. MEAN**	COMP. MEAN	M DIFF.	t	p
1. I would like to attend another meeting such as today's.	2.29	2.36	.07		*
2. I thoroughly enjoyed today's meeting.	2.27	2.28	.01		*
3. I tended to experience conflict in today's meeting.	3.67	4.29	.62	1.63	.20
4. The needs of other members conflicted with my own.	3.64	4.23	.59	1.48	.20
5. I benefited quite a bit from discussing the problem with the group.	4.22	4.30	.08		*
6. I contributed a good deal to the group's work.	3.31	3.67	.37	1.08	.30
7. Our group was pretty successful.	4.00	3.42	.58	2.19	.05
8. My success in the group was pretty dependent on the other members also succeeding.	3.67	4.55	.88	2.35	.05
9. The goal of the group was very important for me.	3.29	3.58	.29	.735	*
10. I was very involved in the group's problem.	3.36	3.10	.26	.686	*
11. I accepted the goal of the group as worthy of my best efforts.	2.29	2.93	.64	1.84	.10
12. I felt it was necessary to agree with the other members about a common solution to the problem.	3.89	4.45	.56	1.51	.20
13. I felt it was more necessary to compete with than cooperate with the other members.	5.06	4.80	.26	.765	*
14. If I get a rating of one it will be due mainly to my own efforts.	3.18	1.86	1.32	4.09	.001
15. If I don't get a rating of one, it will be due mainly to the other people in the group.	4.50	5.36	.86	2.69	.01
16. At times I felt angry toward some of the other members.	4.15	4.20	.05		*
17. Giving information to the other members benefited me.	3.14	2.74	.44	1.18	.30
18. I felt my individual reputation was at stake during the group meeting.	4.98	5.52	.54	2.17	.05

* Not significant.

** A score of 1 indicates strong agreement; 2, moderate agreement; 3, slight agreement; 4, slight disagreement; 5, moderate disagreement; and 6, strong disagreement.

TABLE 3

DIFFERENCES BETWEEN COMPETITIVE AND COOPERATIVE INDIVIDUALS AS SHOWN IN CONTENT ANALYSIS OF WRITTEN COMMUNICATIONS*

CONTENT			CHI SQUARE	p
1. Indicates cohesiveness	does	does not		
Coop.	26	19		
Comp.	14	31	5.49	.02
2. Attempts influence	does	does not		
Coop.	26	19		
Comp.	13	32	6.56	.02
3. Sends pressures toward uniformity	does	does not		
Coop.	19	26		
Comp.	6	39	7.98	.01
4. Sends relevant content (mentions specific solution to problem)	less than 3	3 or more		
Coop.	24	21		
Comp.	34	11	3.90	.05
5. Number of words sent	under 150	150 or more		
Coop.	18	27		
Comp.	27	18	2.84	.10
6. Communications "received"†	6 or less	over 6		
Coop. "sender"	10	26		
Comp. "sender"	31	5	23.57	.001
7. "Receives" expression of tension	0–1	2 and more		
Coop. "sender"	33	3		
Comp. "sender"	22	14	7.56	.01
8. "Receives" expression of antagonism	does	does not		
Coop. "sender"	5	31		
Comp. "sender"	14	22	4.39	.05
9. "Receives" opinion	less than 10	10 or more		
Coop. "sender"	6	30		
Comp. "sender"	28	8	24.21	.001
10. "Receives" information	less than 4	4 or more		
Coop. "sender"	15	21		
Comp. "sender"	30	6	11.62	.001
11. Communicates tension	does	does not		
Coop.	10	35		
Comp.	18	26	2.73	.10
12. Uses nondirective more than directive manner of influence‡	does	does not		
Coop.	16	9		
Comp.	5	8	4.76	.05
13. Takes individual role	does	does not		
Coop.	18	27		
Comp.	23	21	.87	

basis that cooperative *S*s have their goal mediated by the group, while this is not the case for competitive *S*s. The results of this experiment support that prediction (Table 3, item 1). A content analysis of communications found cooperative *S*s showing cohesive behavior (speaking of the group as a whole, using words: *group*, *we*, *us*, *others*, etc.) signifi-

TABLE 3 (*Concluded*)

	observed frequency	expected frequency		
14. "Receives" individual role				
Only to coop. "sender"	2	15		
Only to comp. "sender"	29	15	24.30	.001

* All chi squares contain Yates's correction for continuity.
† There were a total of 72 "receivers" in 18 groups.
‡ Just includes *S*s who made at least one influence attempt.

cantly more so than did competitive *S*s. This finding supports the assumption of cooperation as a determinant of cohesiveness. This measure of cohesiveness is similar to that used by Polansky, Lippitt, and Redl (11).[4]

From the conceptual definition of cohesiveness, a prediction was made that cooperative *S*s would more frequently attempt social influence (an *explicit attempt* to modify the behavior of the recipient, such as accepting a particular solution to the human relations problem). This received substantial support from the results of the experiment. The number of cooperative *S*s who attempted to influence others was significantly greater than the number of competitive *S*s (see Table 3, item 2, $p < .02$). This concurs with Back's finding of cohesiveness as a source of influence (1).

Results in Table 3, item 3, substantiate ($p < .01$) the hypothesis that cooperative *S*s would send more pressures toward uniformity than competitive *S*s (favors reaching *common agreement* by all members). Festinger (4) hypothesizes that high cohesiveness is a source of increased pressure toward uniformity for individuals in a group. Since cooperative *S*s were to be rated as a group, they were subjected to greater pressures toward uniformity than were competitive *S*s.

Festinger states that "pressures toward uniformity will be greater, the more dependent the various members are on the group in order to reach their goals" (4, p. 273). This condition was nonexistent for the competitive *S*s and explicitly defined for cooperative individuals. If an *S* reaches his goal only when others do not (competitive), agreeing with others would deprive him of success.

Responses to the love-punishment scale (Table 4) support the hypothesis that cooperative *S*s are more likely to accept pressures toward uniformity than are competitive *S*s. Festinger and Thibaut (7) also report greater opinion changes when the pressure toward uniformity is increased. Festinger has formulated the following hypothesis with which the results here concur: "*The amount of change in opinion resulting from receiving a communication will increase as the pressure towards uniformity in the group increases* [for the recipient]" (4, p. 276).

[4] The percentage agreement for the categories ranged from 71 (attempts influence) to 97 for mentions specific solutions to the human relations problem (relevance).

TABLE 4

CHANGES IN RESPONSE TO LOVE-PUNISHMENT SCALE SHOWING ACCEPTANCE OF PRESSURE TOWARD UNIFORMITY

GROUP	DIFF. MEANS	*t*	*p*
A. CHANGE FROM FIRST TO SECOND RESPONSE			
Competitives	.103	.791	*
Cooperatives	.278	2.07	.05
B. CHANGE FROM SECOND TO THIRD RESPONSE			
Competitives	.026	.176	*
Cooperatives	.361	3.34	.01

* Not significant.

Deutsch (3) also reports that "Indiv coop [*sic*] were affected by the ideas of others significantly more than were Indiv comp [*sic*]."

As predicted, communications containing relevant information were more frequently sent by recipients of cooperative instructions. Explicit mentions of specific solutions to the problem were considered relevant communications (Table 3, item 4). The relationship was found to be statistically significant at the .05 level of confidence.

When an issue is relevant for group locomotion, there will be more pressures on members to communicate with one another about it (4, 12). Here, mentions of specific solutions to the problem probably were one of the most relevant topics possible. This result is substantiated by Schachter (12), who reports that pressures to communicate are strongest under conditions of relevance to the group.

Cooperation as a Determinant of Instrumental Communications

The results of this experiment show that cooperative *S*s communicated more words

per individual and less tension than did competitive *Ss*. However, the differences (Table 3, items 5 and 11) are not statistically significant ($.05 < p < .10$).

The failure to establish definitely the hypothesis possibly can be explained in part at least by the nature of the experimental design. Since competitive *Ss* as well as cooperative *Ss* were given the alternatives of sending messages or thinking alone the entire hour, and since a norm of communicating by message was introduced with the standard stimulus messages, there were forces on the competitive *Ss* to communicate. They knew that other individuals possibly were writing during the first "note-writing" period and received messages supposedly from the other group members.

There are marked differences between cooperative and competitive *Ss* (owing to the standard stimulus messages) when they are viewed as recipients of communications. Cooperative *Ss* (i.e., perceived sources of set A and set C notes) were much more likely to "receive" communications than were competitive *Ss* (i.e., perceived sources of set B and set D notes). The results reported here substantiate the hypothesis ($p < .001$, Table 3, item 6).

A selective factor seemed to operate that channeled certain communications to specific "recipients." Individuals who "sent" opinions and information "received" such communications from their recipients. The same held true for communications of tension and antagonism. As Festinger, Schachter, and Back report, "communication occurs towards those seen as being most affected by the information" (5, p. 130) Cooperative *Ss* invited communications in the messages they sent, while the competitive standard stimulus notes that were from "B" explicitly restrained communications of any type and those of "D" invited answers not relevant or hostile to the problem at hand.

Cooperative *Ss* were more likely to receive instrumental communications (opinion and information) than were competitive *Ss* ($p < .001$). Cooperative *Ss* were significantly less likely to receive consummatory communications (tension and antagonism) than were competitives (Table 3, items 7, 8, 9, 10). This may be explained in terms of the tendency for instrumental communications to be channeled toward cooperative recipients and for cooperative notes to induce restraining forces against noninstrumental answers.

Of the 21 *Ss* who used nondirective type social influence, 16 were recipients of cooperative instructions. The difference is significant ($p < .05$, Table 3, item 12). Striving for a goal that is mediated by their group makes social communication instrumental for cooperative goal achievement. Cooperative individuals, rated as a group and having a commonality of fate, are more likely to influence one another than are competitives. This basic situation makes it less necessary for cooperatives to exert considerable effort in attempting social influence. They need not expect too much resistance, all other things being equal.

For competitive *Ss*, the situation is quite different. Their goal achievement occurs through individual effort and is facilitated by the failure of others. Social influence is not theoretically necessary for them. Competitive *Ss* can more readily resist inductions, and stronger influence attempts probably would be necessary. Hence, the difference reported concerning nondirective influence could be theoretically expected.

Findings by Deutsch (3) that cooperative *Ss* are more likely to take group functions and less likely to take individual functions than are competitives were not substantiated here (Table 3, item 13). No significant differences were found here between cooperative and competitive individuals in the taking of group and individual membership roles.

One possible explanation is that the hypotheses in this research neglected the situation in which they were tested. Deutsch studied groups in face-to-face interaction. It may be that note-writing interaction is less likely to produce individual role taking. If further research substantiates this finding, an important difference will be found to exist between ordinary small groups and those situations in which communications are restricted to written messages. Festinger (4) suggests that defensive reactions are more commonly aroused in face-to-face groups.

When the notes containing individual roles were further examined and a comparison made between the number of persons who take "individual" roles only to competitive "senders" and those who do so only to cooperative "senders," a statistically significant difference

was found. Twenty-nine persons took individual roles toward just competitives ("senders" B and D), while only one person took an individual role towards a cooperative "sender" (Table 3, item 14). The same selective factor mentioned in explanation of differences in the receipt of consummatory and instrumental communications possibly was in operation here.

Conflicts in Expectations

Since the planted notes were heterogeneous, they were bound to conflict in one way or another with the expectations of the recipients as structured by the experimental instructions. Some unanticipated questionnaire results suggest that the *frame of reference of an individual will determine his expectations of others*. From this, two additional hypotheses may be inferred: (*a*) Individuals who perceive themselves as cooperative will tend to expect cooperative behavior from one another, and (*b*) individuals who perceive themselves as competitive will tend to expect competitive behavior from one another.

Competitive *S*s tended to accept the item "Our group was pretty successful" (Table 2) significantly more often than did cooperative *S*s. This result may be related to a difference in the frame of reference and expectations held by cooperative and competitive *S*s. Competitive *S*s could legitimately expect no help from other people in the group. Yet they received some. On the other hand, cooperative *S*s could expect, on the basis of the instructions they received, cooperative behavior from all the other group members. This they did not receive. Competitive expectations probably were exceeded by their receiving help from cooperative individuals, while the expectations of those receiving cooperative instructions were not met during the group meetings.

Although not statistically significant, there was a tendency for cooperative *S*s to agree more often than competitive *S*s (Table 2) that "I tended to experience conflict in today's meeting" and "The needs of other members conflicted with my own." Also, cooperative *S*s were less likely to agree that "Giving information to the other members benefited me." These findings suggest the same interpretation. A detailed analysis of sociometric responses in terms of conflicts in expectations is presented in (10).

SUMMARY AND CONCLUSIONS

From previous research on cooperation, competition, and studies of small groups, an attempt was made to formulate hypotheses that would integrate theoretical approaches to both areas of investigation. The main problem selected for study was to determine the consequences of cooperation and competition on small group cohesiveness, social influence, and communication.

The results of the experiment permit the following conclusions:

1. Cooperation may be considered a determinant of group cohesiveness. Cooperative *S*s showed significantly more cohesive behavior, attempts at influence, exertion and acceptance of pressures toward uniformity, and communications of relevance than did competitive *S*s.

2. Cooperation may be considered a determinant of instrumental communications. Cooperative *S*s (attributed senders of cooperative messages) received significantly more communications, more instrumental communications (opinion and information), and fewer consummatory communications (tension and antagonism). Also, they tended to send more communications and fewer consummatory ones (showing tension), but these findings were not quite significant statistically. A significantly greater frequency of nondirective influence by cooperatives was explained in terms of the fact that considerable effort in social influence is theoretically not necessary for cooperative *S*s as compared to competitive *S*s.

A prediction that cooperative *S*s would take more group and fewer individual functions than competitive *S*s was not supported. This was explained as possibly due to the fewer defensive reactions reported for note-writing group situations.

The experimental results support the following *ad hoc* hypotheses:

1. The frame of reference of an individual will determine his expectations of others.

a. Individuals who perceive themselves as cooperative will tend to expect cooperative behavior from one another.

b. Individuals who perceive themselves as competitive will tend to expect competitive behavior from one another.

Much remains unknown concerning differences between face-to-face and note-writing

groups. In problem-solving settings, the latter may have much to offer as far as efficiency for research purposes is concerned.

The extent to which the norm introduced by having all *S*s receive communications and having the opportunity to send them caused competitives to communicate possibly can be determined by subsequent study. Also, the extent to which cooperative messages induced restraining forces against competitive and out-of-field responses should be determined.

Further research is needed to study the development of channels of communication in groups. The differences between cooperative senders in the content of communications received suggests some things about "who gets what" in the communication process. It would seem valuable to study perceptions that senders have of possible recipients and determinants of the content sent to specific receivers.

The relationship between determinants of cohesiveness and instrumental communication has not been studied intensively. Possibly, in situations where cohesiveness is due to attractiveness of members or group prestige, one may not expect as much instrumental communication as under conditions of task direction.

REFERENCES

1. BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–24.
2. DEUTSCH, M. A theory of cooperation and competition. *Hum. Relat.*, 1949, **2**, 129–152.
3. DEUTSCH, M. An experimental study of the effects of cooperation and competition upon group process. *Hum. Relat.*, 1949, **2**, 199–232.
4. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.
5. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups.* New York: Harper, 1950.
6. FESTINGER, L., SCHACHTER, S., BACK, K., KELLEY, H., & THIBAUT, J. *Theory and experiment in social communication.* Ann Arbor: Univer. of Michigan, Research Center for Group Dynamics, 1950.
7. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.
8. GROSSACK, M. The effect of cooperation and competition upon small group cohesiveness, social influence and communication. Unpublished doctor's dissertation, Boston Univer., 1952.
9. GROSSACK, M. Controlling interaction in small group research. *J. Psychol.*, 1953, **35**, 241–244.
10. GROSSACK, M. Cues, expectations and first impressions. *J. Psychol.*, 1953, **35**, 245–252.
11. POLANSKY, N., LIPPITT, R., & REDL, F. An investigation of behavioral contagion in groups. *Hum. Relat.*, 1950, **3**, 319–348.
12. SCHACHTER, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.

Received March 13, 1953.

ETHNOCENTRISM AND TOLERANCE OF TRAIT "INCONSISTENCY"

IVAN D. STEINER

University of Illinois

PERCEPTION and memory are "selective" processes in which some attributes of complex stimulus objects become accentuated while others are ignored or denied (2, 3). When an individual assumes that certain attributes belong together, this assumption tends to determine which attributes of a complex stimulus object he will perceive and remember (2, 3, 5, 8). These two propositions provide the basis for an approach to problems of interpersonal perception in social relations.

The behaviors of interacting persons may be presumed to be influenced by the percepts they hold of one another. According to the above propositions, these percepts are formed and maintained under the guiding influence exerted by the perceiver's assumption that certain human attributes belong together while others do not. Thus, the individual's assumption that certain attributes belong together is expected to influence his percept of the person with whom he is interacting; the percept is expected to influence his behavior toward that other person. From a research point of view, this theoretical orientation suggests that knowledge of an individual's perceptual assumptions should enable one to make predictions concerning his percepts and behaviors. Conversely, it should be possible to predict the nature of an individual's perceptual assumptions from a knowledge of his percepts or behaviors. If two categories of subjects are known to hold different percepts, it may be inferred that their percepts are based on different sets of perceptual assumptions. Given a knowledge of the differences between the percepts which are held by these two categories, it should be possible to make predictions concerning differences between the perceptual assumptions they are employing. This paper reports a study in which predictions of this kind are made and tested. It investigates the perceptual assumptions made by two types of people—types which have been shown by other researchers to hold different percepts of the persons with whom they interact.

Persons with a high degree of ethnocentric prejudice and persons with a low degree of ethnocentric prejudice were chosen to represent two distinct types of perceivers. Two major considerations led to this choice. There is empirical evidence indicating that ethnocentrism is associated with differences in cognition. Studies such as those reported by Adorno *et al.* (1) have found that high ethnocentrics differ considerably from low ethnocentrics in their percepts of people. Ethnocentrism is positively associated with perceptual perseveration (4) and "intolerance of ambiguity" (4, 6, 7). In addition to providing types of perceivers who hold different percepts of people, the ethnocentrism variable is one in which social psychologists have had a practical interest. For both of these reasons, high and low ethnocentrics were selected for investigation. However, it should be apparent that this choice was somewhat arbitrary. Whenever *any* two types of perceivers hold different percepts, it seems pertinent to inquire whether this difference may be a consequence of their holding different perceptual assumptions.

This study not only attempts to determine whether high and low ethnocentrics differ from one another in their perceptual assumptions; it also formulates certain postulates concerning the systematic nature of the difference between the perceptual assumptions of high and low ethnocentrics. Two hypotheses will be derived and tested.

THEORETICAL FORMULATION

Definitions of Terms

Trait contingency. The problem is one of comparing the assumptions which high and low ethnocentrics make concerning the relationship of one human trait to another. If high ethnocentrics assume that two designated attributes are likely to occur together, will low ethnocentrics also expect these two attributes to occur together? In the following discussion the term *trait contingency* will be used to denote the level of probability which an individual perceiver attributes to the co-occurrence of two traits. If he thinks a given

pair of traits have a high probability of occurring together, those two traits have a high trait contingency for him. If he feels that they have a low probability of occurring together, they have a low trait contingency for him.

Trait discrepancy. All of us are accustomed to making evaluative judgments of human traits. We say that honesty and intelligence are "good" traits and that laziness and unfaithfulness are "bad" traits. We would probably experience little difficulty in assigning most traits to positions along a continuum representing various degrees of "goodness" vs. "badness." In the following discussion the term *trait discrepancy* will be used to denote the distance separating two traits along an evaluative dimension, the ends of which are defined by the polar terms good and bad. If two traits are separated by a short distance (i.e., have very similar evaluative connotations), they have a small trait discrepancy. If two traits are separated by a long distance (i.e., have very dissimilar evaluative connotations), they have a large trait discrepancy.

Postulates

It may be expected that the trait contingency of any pair of traits will be partially determined by the degree to which the two traits are alike on certain dimensions of meaning. If two traits connote about the same degree of activity vs. passivity, or the same degree of goodness vs. badness, they may, for that reason, have a high trait contingency. This study concentrates on one dimension of trait similarity: the evaluative, or good-bad dimension. It investigates the relationship between trait contingencies and trait discrepancies, and attempts to determine whether the relationship between these two variables is greater for high ethnocentrics than it is for low ethnocentrics.

The most general assumption concerning the relationship between these variables may be stated as follows: trait contingencies will vary as an inverse monotonic function of trait discrepancies. This assertion contends that the larger the trait discrepancy (other things being equal), the smaller will be the trait contingency. Any unit of trait discrepancy will have the effect of lowering the trait contingency of a pair of traits. Many units of trait discrepancy will reduce trait contingencies by a larger amount than will few units of trait discrepancy.

The above statement of relationship makes no allowance for a differential between high and low ethnocentrics. Studies such as those reported by Adorno *et al.* (1) would seem to suggest that high ethnocentrics are more sensitive than are low ethnocentrics to differences along a good-bad dimension. For example, Frenkel-Brunswik (1, ch. 11) described her low-ethnocentric *S*s as more likely to see a heterogeneous mixture of good and bad traits within any one person. They seemed more likely than high ethnocentrics to "tolerate" the co-occurrence of pairs of traits which have large trait discrepancies. Therefore, it seems desirable to refine the above statement of the relationship between trait contingencies and trait discrepancies.

Let us propose that high and low ethnocentrics do, in fact, differ in sensitivity to trait discrepancies. High ethnocentrics, we shall say, are more sensitive to the trait discrepancy factor. Any unit of trait discrepancy will have a larger effect on the trait contingencies of high ethnocentrics.

Postulate 1. The effect of any unit of trait discrepancy separating one trait from another will be to decrease the trait contingency of those traits, the amount of this decrease being larger for high ethnocentrics than for low ethnocentrics.

Figure 1 presents this postulate graphically. As trait discrepancies become progressively larger (i.e., as more and more units of trait discrepancy are added), trait contingencies

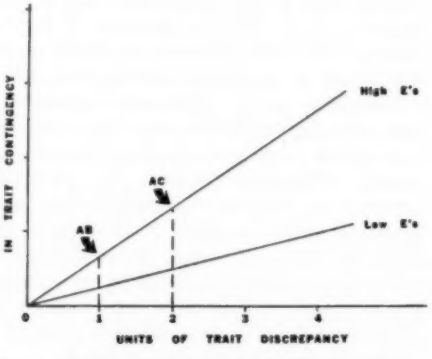

FIG. 1. POSTULATED EFFECT OF THE TRAIT DISCREPANCY FACTOR ON TRAIT CONTINGENCIES

sustain increasing amounts of total decrease. Furthermore, this total amount of decrease is always larger for high ethnocentrics than for low ethnocentrics. Every unit of trait discrepancy is shown to result in a larger decrease in the trait contingencies of high ethnocentrics than in the trait contingencies of low ethnocentrics. (It is to be noted that the two lines need not be straight in order to satisfy the requirements of this postulate.)

Postulate 1 has dealt exclusively with the possible effects of the trait discrepancy factor. It is now necessary to consider the effects which other factors may have on trait contingencies.

Postulate 2. The combined effects of all other factors will not be such that they cancel the differences between the trait contingencies of high and low ethnocentrics which have been created by the trait discrepancy factor.

It is possible that certain other factors may have differentiating effects on the trait contingencies of high and low ethnocentrics. If these differentiating effects are negatively correlated with those exerted by the trait discrepancy factor, they would tend to cancel the differentiating effect of the latter factor. The two lines shown in Fig. 1 would then assume the same slope; the uncontrolled effect of other factors would eliminate the difference in slope which would have been present if the trait discrepancy factor had been the only determinant of trait contingencies.

This postulate does not deny that other factors may be potent determinants of trait contingencies. It denies only that such other factors will have a combined effect which systematically differentiates between high and low ethnocentrics in a direction opposite to the differentiating effect of the trait discrepancy factor. In substance, it asserts that the relationship shown in Fig. 1 will prevail even when other factors are permitted to vary.

Hypotheses

Consider the case in which a trait, A, is separated from another trait, B, by one unit of trait discrepancy. The presence of this unit of trait discrepancy will decrease the trait contingencies which both high and low ethnocentrics will give to this pair of traits. This fact is indicated in Fig. 1 by the positive slope of both lines. However, the steeper slope of the line representing the relationship for high ethnocentrics indicates that the presence of this one unit of trait discrepancy will cause a larger decrease in the trait contingency which high ethnocentrics give to this AB pair. The trait contingency which low ethnocentrics give to this AB pair will be relatively less affected by the presence of this unit of trait discrepancy. Consider now the case in which trait A is separated from a third trait, C, by two units of trait discrepancy. The trait contingencies which high and low ethnocentrics will give to this AC pair will be even more *differentially* decreased than was true in the case of the AB pair. The difference between the amounts of depressive influence exerted by the trait discrepancy factor on such AB and AC pairs will always be greater for high ethnocentrics.

If, in the above case, the AC pair had had a trait discrepancy of three units, the difference between high and low ethnocentrics would have been even more striking. In the case of high ethnocentrics, the AB and AC pairs would have sustained markedly different amounts of decrease. But in the case of low ethnocentrics, these two pairs would have sustained only moderately different amounts of decrease. Given the steeper slope of the line representing the relationship for high ethnocentrics, we may generalize as follows: the greater the difference between the trait discrepancies of an AB and AC pair, the more dissimilar high and low ethnocentrics will be with respect to the amount by which the decrease sustained by the AC pair exceeds that sustained by the AB pair.

Let us confront both high and low ethnocentrics with a series of test items. Each item will consist of an AB and an AC pair of traits; the AC pair will always have a larger trait discrepancy than the AB pair. The task of *S* is to cross out the pair in each item which has the lower trait contingency (i.e., the pair which consists of traits having the lower probability of co-occurrence). What hypotheses can be advanced concerning the responses of high and low ethnocentrics in such a test situation?

It cannot be hypothesized that all *S*s, regardless of ethnocentrism, will tend to cross out the AC pairs. Factors other than the trait discrepancy factor might combine to give the AC pairs very high trait contingen-

cies. Even though the trait discrepancy factor may decrease the trait contingency of the AC pair by a considerable amount, it is still very possible that the resultant trait contingency for the AC pair may be higher than that for the AB pair. In the absence of controls over such other factors, no hypothesis could be made concerning the comparative trait contingencies of the AB and AC pairs.

However, it is possible to advance hypotheses concerning the *differences* between the responses of high and low ethnocentrics. Regardless of the effect of other determining factors, the trait discrepancy factor will exert a more marked effect on the trait contingencies of high ethnocentrics. In the case of high ethnocentrics, it will exert a stronger depressive influence on the AC pair than on the AB pair. In the case of low ethnocentrics, it will exert more nearly equal influences on the AB and AC pairs. Consequently, the amount of decrease sustained by the AC pair relative to the amount sustained by the AB pair should tend to be greater for high ethnocentrics. This means that the resultant trait contingency for the AC pair relative to that of the AB pair should be smaller for high ethnocentrics than for low ethnocentrics.

Therefore, given test items in which the AC pair always has a larger trait discrepancy than the AB pair, we may predict:

Hypothesis 1. High ethnocentrics will cross out more AC pairs than will low ethnocentrics.

We have already noted that when the AB and AC pairs have markedly different trait discrepancies, the trait discrepancy factor should distinguish more clearly between high and low ethnocentrics. The greater the difference between the trait discrepancies of the AB and AC pairs, the more dissimilar should be the trait contingencies given by high and low ethnocentrics.

Hypothesis 2. The greater the difference between the trait discrepancies of the AB and AC pairs in any test item, the greater will be the excess of high ethnocentrics over low ethnocentrics who will cross out the AC pair.

METHOD

In a preliminary data-gathering step, 30 Florida State University students rated the desirability of 32 traits. A ten-point rating scale was used. A computation was then made of the mean rating received by each trait. With these mean ratings as criteria of evaluative connotation, 18 test items were prepared. Each test item consisted of a selected trait A paired successively with a trait B and a trait C. Different traits were used in different test items, but in every case the selection of traits was such that traits A and B had received approximately equal mean ratings on the evaluative continuum. Trait C was always a trait which had received a somewhat different evaluative placement. Randomization was employed in determining whether the AB or AC pair would occupy the left-hand position in a test item. The first test item was: (passionate—quick-tempered) (lazy—quick-tempered).[1]

Fifty-two Florida State University students responded to the test items by crossing out the pair of traits in each item which seemed less likely to occur together.[2] The tests were scored by counting the number of items to which the *S* responded by crossing out the AC pair (i.e., the pair with the larger trait discrepancy). Each *S* also responded to the California Ethnocentrism Scale (1, p. 142) which will be referred to here as the E scale. The *S*s also reported their academic averages at the university.

In order to obtain a more precise estimate of the trait discrepancies of the AB and AC pairs, a second group of 31 students was asked to rate the traits on two continua designed to represent the evaluative dimension. These two continua were defined by the polar terms: good-bad and worthless-valuable.[3] The 31 judges indicated the position of each trait on each continuum by placing a check mark on a seven-point graphic rating scale, a technique copied from Osgood's[4] Semantic Differential. In computing the judge's rating of each trait, the seven positions along the two continua were assigned numerical values ranging from 1 (for the bad or worthless end of the continuum) to 7. The mean of the ratings given to each trait on these two continua was taken to represent the individual judge's rating of the trait. The mean of the ratings given to each trait by

[1] This instrument and a table showing the trait ratings and trait discrepancies of the test items used in the instrument have been deposited with the American Documentation Institute. Order Document No. 3983 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[2] The writer is indebted to Basil Zimmer of Florida State University for cooperation in gathering these data.

[3] These two continua were selected to represent the evaluative dimension because Osgood (see footnote 4) had previously found them to have factor loadings of .88 and .79 on what appeared to be an evaluative factor. The judges also rated each trait against two continua representing the "activity" dimension (active-passive and slow-fast) and two continua representing the "potency" dimension (strong-weak and small-large). Because the trait ratings on these two dimensions correlated highly (rho = .80 and .92) with ratings on the evaluative dimension, the activity and potency dimensions will not be treated here as independent dimensions.

[4] Osgood, C. E. Personal communication, December, 1952.

all 31 judges was used to represent the position of the trait along the evaluative dimension. (The standard deviations of the judges' ratings ranged from .70 to 1.36.) When these mean ratings were assigned to the traits used in the 18 test items, each item was found to conform to the criterion; in each item the AB trait discrepancy was smaller than the AC trait discrepancy.

RESULTS

As a test of hypothesis 1, *S*s whose E scores fell in the top quarter of the distribution of E scores were compared with *S*s whose E scores fell in the bottom quarter of the distribution. The *S*s whose E scores fell in the top quarter crossed out a mean of 8.69 AC pairs. The *S*s whose E scores fell in the bottom quarter crossed out a mean of 6.85 AC pairs ($t = 2.15$). Because both of the hypotheses tested in this research were formulated before the data were gathered, only one tail of the t distribution was used in computing probabilities. The probability corresponding to a t of 2.15 is less than .03. High E's crossed out significantly more AC pairs than did low E's.

Hypothesis 2 was tested by correlating the amount of difference between the trait discrepancies of the two pairs in each item with the numerical excess of high E's over low E's who crossed out the AC pair in that item. This product-moment correlation was +.57 ($p < .01$). Items in which there is a large difference between the AB and AC trait discrepancies tend to discriminate best between high and low ethnocentrics.

DISCUSSION

Confirmation of these two hypotheses offers support for the contention that high and low ethnocentrics do differ in the assumptions they make concerning the co-occurrence of human traits. Furthermore, the differences between the assumptions made by high and low ethnocentrics appear to be systematically related to a difference in sensitivity to trait discrepancies. High ethnocentrics tolerate less disharmony between their value systems and their perceptual assumptions.

It is necessary to consider certain alternative explanations of the data. The test items were similar in form to items included in many intelligence tests. Adorno *et al.* (1, ch. 8), and O'Connor (6) have obtained moderate negative correlations between indicators of intelligence and E-scale scores. Therefore, it is possible that the greater tendency of high E's to cross out the AC pairs is somehow a function of their lower intelligence. However, in the present sample, the rank-order correlation between college grades, as indicators of intelligence, and the number of AC pairs crossed out was only +.05. Another test of this alternative explanation of the data was provided by a comparison of the mean number of AC pairs crossed out by students with high, and with low, academic averages. The 16 students with academic averages of B− or above crossed out a mean of 8.38 AC pairs, and the 12 students with averages of C− or below crossed out a mean of 8.58 AC pairs. The t value for this difference between means is .26 ($p > .70$). In view of these findings an explanation based on the intelligence factor seems untenable.

Another explanation would hold that the high E's differ from low E's in the evaluative connotations they give to the trait names used in this study. If this were so, the differences between the trait discrepancies of the AB and AC pairs might not be the same for high E's as they are for low E's. High E's might then have crossed out more AC pairs simply because the differences between the AB and AC trait discrepancies were larger for them than they were for low E's. In order to test this possibility, the mean trait ratings given by high-E judges were computed separately from those given by low-E judges. The difference between the trait discrepancies of the AB and AC pairs in each item was then computed under two conditions: (*a*) the traits carry the mean rating assigned by high-E judges (Condition 1), and (*b*) the traits carry the mean ratings assigned by low-E judges (Condition 2). For the 18 test items, the mean difference between the trait discrepancies of the AB and AC pairs was only .09 larger under Condition 1 than under Condition 2 ($p > .80$). As a further test of this alternative explanation, a computation was made of the amount by which the difference between the trait discrepancies of the pairs in each item was larger under Condition 1 than under Condition 2. These indexes were computed separately for each item and then correlated with the previously described measure of the ability of the item to discriminate between high- and low-E scorers (i.e., number of high E's who crossed

out the AC pair minus the number of low E's who crossed out the AC pair). The obtained product-moment correlation was $-.22$ ($p > .30$). Not only does this correlation fail to reach statistical significance, but it is in the direction opposite to that demanded by this alternative explanation. In view of these findings, this explanation seems untenable.

SUMMARY

This study has investigated the assumptions which two kinds of people make concerning the likelihood that various human traits will occur together. Interest in the problem derives from a theoretical orientation which maintains that such assumptions lead to selective perceptions and memories, and hence exert an effect on the percepts which an individual forms of other people. If this orientation is sound, knowledge of an individual's percepts should permit predictions about the assumptions which he employs in forming concepts.

Because other researchers have found that high and low ethnocentrics differ in their percepts of people, it was inferred in this study that they also differ in the assumptions they make concerning the interrelationships of human traits. It was postulated that when two human traits are considered to be about equally desirable, high ethnocentrics will assume that these traits have a high probability of occurring together. If two traits are unequal in desirability, high ethnocentrics are expected to assume that these traits are unlikely to occur together in other people. In the case of low ethnocentrics, it was postulated that the relative desirability of two traits would have less effect on the assumed probability of their co-occurrence.

Two hypotheses were derived and tested. As a result of these tests, the null forms of the hypotheses were rejected at the .03 and .01 levels. High and low ethnocentrics do make different assumptions concerning the likelihood that various pairs of human traits will co-occur. Furthermore, the differences between the assumptions made by these two kinds of perceivers are in the predicted direction: the assumptions which high ethnocentrics make are more closely related to their appraisals of the desirability of the traits.

Although the data in this study have dealt exclusively with differences between the perceptual assumptions made by high and low ethnocentric perceivers, the problem under investigation has a wider scope. In its broader implications, the problem is one of determining the effect which perceptual assumptions may have upon percepts, and ultimately upon the behavior of interacting persons. The present study has been successful in using other researchers' findings concerning percepts as a basis for predicting perceptual assumptions. Additional research is needed to determine the extent to which percepts and behaviors are a consequence of perceptual assumptions.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ALLPORT, G. W., & POSTMAN, L. J. The basic psychology of rumor. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Henry Holt, 1947. Pp. 547–558.
3. BARTLETT, F. C. *Remembering*. Cambridge, Eng.: Cambridge Univer. Press, 1932.
4. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
5. HOROWITZ, E. L., & HOROWITZ, R. E. Development of social attitudes in children. *Sociometry*, 1939, **1**, 301–338.
6. O'CONNOR, P. Ethnocentrism, "intolerance of ambiguity" and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526–530.
7. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
8. ZILLIG, M. Einstellung und Aussage. *Z. Psychol.*, 1928, **106**, 58–106.

Received March 23, 1953.

THE EFFECT OF STUTTERING ON THE BEHAVIOR OF THE LISTENER[1]

SEYMOUR ROSENBERG AND JAMES CURTISS[2]

Human Resources Research Center, Randolph Air Force Base, Texas

Pennsylvania State University

THE severity of stuttering behavior has been shown to vary as a function of a number of conditions. In a recent review of the literature, Bloodstein (2) has attempted to order the rather heterogeneous list of conditions which have been found to affect stuttering. He classified these conditions into six general types. One of his descriptive classes of interest for the present study is the behavior of the listener. Studies summarized under this class are predominantly concerned with the status of the listener as a friend, stranger, member of the family, etc., rather than the behavior of the listener during stuttering. Bloodstein noted, "The kinds of listener behavior which tend to facilitate the stutterer's speech do not appear to be specifically defined in the literature" (2, p. 296). Reports obtained from stutterers by personal account and by investigators (3, 6, 7) seem to indicate that the "stutterer speaks best when there seems to be no reaction to his stuttering at all" (2, p. 297).

But what *are* the listener reactions which are peculiar to stuttering? Before attempting an experimental answer to the problem of the effect of the behavior of the listener on the severity of stuttering, we must have rather clear descriptions of what the listener does. Descriptions by the stutterer of listener behavior are complicated by the fact that the stutterer is so intimately involved in the speaking situation that his report may differ considerably from an objective analysis of the speech situation.

A review of the literature reflects a paucity of objective descriptions of listener reactions to stuttering speech. A study by Ainsworth (1) may be considered as somewhat related to this problem. He investigated the breathing of listeners exposed to stuttering. His listeners included normal speakers and stutterers. The experiment was apparently not designed to throw light directly on the problem posed here in view of the exclusion of normal speech as a stimulus condition. A more relevant study is reported by Van Riper (10). He studied the effect of "hearing and seeing" stuttering upon stutterers' speech. Normal speech was included as a control condition. Results indicated that the subjects demonstrated more interruptions following the stuttering condition. All subjects were stutterers, however.

The present experiment is an objective study of some of the listener responses which may be influenced by stuttering. The listeners were all normal speakers. Listener responses chosen by the authors to be studied were (*a*) eye contact, (*b*) hand movement, and (*c*) other bodily movement during "stuttering"[3] and during normal speech.

METHOD

Subjects. Twenty college students from an introductory psychology class at Indiana University were used as *S*s in the study. The *S*s signed up voluntarily although they were required to serve as *S*s in at least three psychological experiments as part of the course requirements. There were 13 males and 7 females with an age range from 17.0 years to 40.5 years, mean age 24.3 years. The study was defined for them as a psychological experiment, with special care being taken to give no cues of the fact that their behavior was being observed in a speech situation. The experiment was performed in the Psychological Clinic.

Procedure. At the time of the appointment, one of the persons who also served as one of the *O*s met the *S* and escorted him to a "waiting room." The *S* was told, "Will you please be seated in here? We will call you in a few minutes. Please stay in the room because we are behind schedule." One of the *E*s was already seated in the room and was posed as an *S* who was also presumably waiting to be called for the experiment. To this *E*, *O* said briefly, "We'll be ready for you in a few minutes." The door was then closed. The *E* started a conversation with *S*. At the end of five minutes of

[1] We wish to express our appreciation to Professor Robert L. Milisen for his suggestions and criticisms in the preparation and performance of this study and to thank Gerald Waite and Norman Willy for their participation as experimenters. The authors assume full responsibility for the statements and conclusions in the paper.

[2] Both authors were formerly at Indiana University where the study was performed.

[3] The "stutterers" were normal speakers who faked stuttering during the experiment. Thus variability in frequency, duration, and severity could be more accurately controlled than would be the case if the *E*s had been actual stutterers.

TABLE 1
SUMMARY OF THE PROCEDURE

SUBJECTS	FIRST FIVE MINUTES	SECOND FIVE MINUTES
Nos. 1, 5, 7, 10, 12	*E* #1, Stuttering	*E* #2, Normal
Nos. 2, 11, 13, 14, 18	*E* #1, Normal	*E* #2, Stuttering
Nos. 4, 8, 9, 15, 17	*E* #2, Stuttering	*E* #1, Normal
Nos. 3, 6, 16, 19, 20	*E* #2, Normal	*E* #1, Stuttering

observation of the *S* through a half-silvered mirror, *O* returned and said to *E*, "We are ready for you now." To *S* he said, "We will be ready for you in a few minutes." As *O* started to close the door, a second *E* walked up and inquired about being an *S* for a psychology experiment. The *O* said to this second *E*, "Will you come in here and wait a few minutes?" The door was closed and a second five-minute observation was begun during which the second *E* talked to the same *S*. Finally *O* re-entered and said to *S*, "O.K., will you come with me?"

The *S* was taken to another room where the "psychology experiment" was begun. This was a brief period in which various things were done with different *S*s as part of pilot studies for other experiments or simply to disguise the fact that the experiment had actually been performed during the so-called "waiting" period. This was done to maintain the naiveté of future *S*s.

The behavior of each *S* was observed under two conditions. These conditions were "stuttering" and normal speech. The order of the two conditions and of the two *E*s was counterbalanced. Thus each *E* stuttered the first five minutes with five *S*s, stuttered the second five minutes with five other *S*s, spoke normally the first five minutes with a third set of five *S*s, and spoke normally the second five minutes with a fourth set of five *S*s. With these restrictions imposed, each *S* was randomly assigned to one of the four possible combinations.

Table 1 is a summary of the counterbalancing procedure. The *S*s shown in the table were numbered in the order in which they participated in the experiment.

Observation and recording of behavior. The waiting room was approximately 7 x 7.5 feet and was arranged so that two comfortable chairs were placed in opposite corners of the room as indicated in Fig. 1. The window (W) in the wall to the right of the *S*s' chair looked out on a path in front of the building with a large area of trees beyond the path. The half-silvered mirror (M) was on the adjacent wall almost directly in front of the *S*.

Two *O*s looked at the *S* through the mirror from a dark room next to the "waiting" room. Each *O* was provided with three electrical leaf switches mounted on a board in front of them. The switches were operated by the *O*s' fingers and corresponded to (*a*) loss of eye

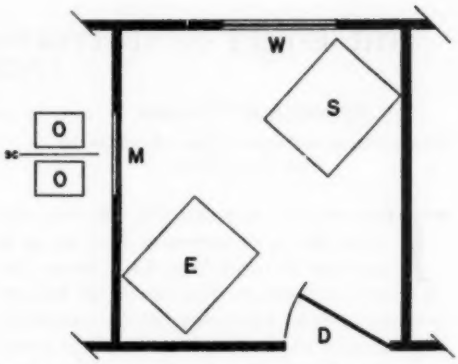

FIG. 1. SKETCH OF THE "WAITING" ROOM AND ADJACENT OBSERVATION ROOM

contact from the *E*, (*b*) hand and/or arm movement, and (*c*) other body movement. These circuit makes and breaks were recorded on an Esterline-Angus Operations Recorder. In order to insure that the *O*s would not respond to each other's judgments, a screen (Sc) was placed between the two sets of leaf switches. As an additional precaution against one *O* being influenced by the judgments of the second, the recorder was placed in a third room so as to minimize the cues provided by the relay clicks. A final precaution was taken to prevent the *O*s from hearing the clicks and/or hearing the conversation in the waiting room by a masking tone delivered through head sets provided for each *O*. Neither *O* knew which condition was being executed at any time since they had no schedule of the random assignment, nor could they see most of the *E*'s body. Figure 1 indicates how part of the mirror was blocked out in the observation room but not in the waiting room.

The conversation was recorded on a tape recorder placed in the *O*s' room with a hidden microphone in the waiting room.

"Stuttering." An attempt was made previous to the experiment to standardize the type and frequency of stuttering and the material discussed in the speaking situation. The stuttering was faked by two graduate students in speech. The stuttering was repetitious and an attempt was made to keep head and body movements to a minimum during a block. Eye contact was maintained during the interruption. There was an attempt by both *E*s to maintain eye contact at a constant frequency and duration during the "stuttering" and "normal" conditions. The first interruption was a severe one and occurred on the first word spoken after *E* and *S* met in the room. A total of 15 interruptions was faked throughout the five-minute period. In order to control as much as possible for conversational content in the situation, a list of topics was selected by the *E*s which included the following: the weather, current events, sports, school major, and "what do you think the experiment will be about?" These topics occurred fortuitously in the two conditions since the *E*s did not have complete control over the topics selected by a particular *S*.

Responses. The two *O*s practiced before the experi-

ment to a point where fair agreement as to what constituted a response was reached. Judgments of each response class were based on the following definitions:

1. "Eye contact loss" was defined as an eye movement (which may or may not have included the head) in any direction away from the *E*'s approximate head position in the room. The total frequency with which an eye contact loss was initiated and *total duration* of these losses in seconds in each five-minute period were obtained from the operations recorder for each *O*.

2. "Hand movement" was defined as any movement of the arm and/or hand excluding single finger tappings. (Single finger movements were usually of such small magnitude as to produce wide disagreement between observers. This was found prior to the experiment and it was decided to omit the recording of such movements entirely.) Total frequency with which a hand movement was initiated and total duration of movement were obtained as measurements from each *O* in each condition. (Smoking was discouraged in the waiting room with a large "No Smoking" sign.)

3. "Other bodily movement" included any observable movement of the body other than head, eye, hand, and arm movements. Again, frequency and duration measurements were obtained.

4. "Speech." After the experiment was completed, both *O*s listened to the tape recordings and measured the total speech time of both *E* and *S* together in each five-minute period.[4] An electric chronoscope, operated by a telegraph key, was used by each *O*.

RESULTS

The total length of time (in seconds) with which each response occurred during a five-minute period was obtained for each *S*. The mean and standard deviation of each response based on twenty *S*s are shown in Table 2. Separate values are shown for each condition and for each observer. The values in the table show the mean length of time that the *S*s (*a*) looked away from the *E*, (*b*) performed hand movements, and (*c*) performed other bodily movements. The response class "total speech of *E* and *S*" refers to the total time spent by both *E* and *S* in talking to each other during the five-minute period.

Table 3 is a summary of the frequency data for each response. The measurement obtained for each *S* is the total number of times each of the three responses was initiated during a five-minute period. For example, "number of eye contact losses" refers to the total number of times an *O* noted that an *S* looked away from the *E* during a particular speech condition.

The seven response measurements presented in Tables 2 and 3 were subjected to similar statistical designs. Separate analyses of variance were computed for each response. The data were first classified according to three variables, i.e., subjects, conditions, and observers. Thus, each cell in such a design contained one measurement of a particular *S* under one of two conditions by one of the two *O*s. For each response, variance estimates were obtained for the three variables, the first-order interactions and the second-order (triple) interactions. With these variance estimates, the reliability of *O*s as well as differences between conditions were evaluated.

Reliability. Several methods of estimating *O* agreement are available from the statistical design. For example, we may test the difference between *O* means. This type of reliability simply requires that the average judgments of each *O* be the same. It would permit considerable inconsistencies between the *O*s from one

TABLE 2

MEAN AND *SD* OF TOTAL DURATION (SECONDS) OF FOUR RESPONSE CLASSES UNDER EACH CONDITION FROM EACH OBSERVER

(N = 20)

RESPONSE CLASS	STATISTIC	OBSERVER #1		OBSERVER #2	
		NORMAL	STUTTERING	NORMAL	STUTTERING
Loss of eye contact by *S*	Mean	91.43	105.20	95.40	110.05
	SD	46.28	58.47	42.13	55.05
Hand movement of *S*	Mean	126.18	126.55	141.10	138.75
	SD	70.78	67.28	66.32	62.05
Other body movement of *S*	Mean	17.65	6.43	25.15	15.35
	SD	21.51	9.65	27.78	15.98
Total speech of *E* and *S*	Mean	270.30	247.54	273.21	252.29
	SD	18.83	24.72	17.17	22.86

TABLE 3

MEAN AND *SD* OF TOTAL FREQUENCY OF THREE RESPONSE CLASSES UNDER EACH CONDITION FROM EACH OBSERVER

(N = 20)

RESPONSE CLASS	STATISTIC	OBSERVER #1		OBSERVER #2	
		NORMAL	STUTTERING	NORMAL	STUTTERING
No. of eye contact losses	Mean	27.85	23.85	23.50	20.95
	SD	8.97	7.54	7.55	9.35
No. of initiations of hand movements	Mean	21.95	18.55	18.85	17.15
	SD	7.72	7.10	6.09	7.32
No. of initiations of other body movements	Mean	8.30	4.60	7.50	5.65
	SD	8.02	5.19	7.55	5.22

[4] Originally, the authors had intended to measure speech time of the *E* and the *S* separately as well as word and sentence completion by the *S*. Owing to the poor quality of the recordings, this was not possible.

S to the next as long as these inconsistencies counteracted each other. A more conservative requirement of reliability is that the *O*s provide the same *relative* measurement for the *S*s. The latter type of reliability allows for a difference between *O*s as long as this difference is consistent for all *S*s and/or for all conditions. The most conservative reliability requirement is a combination of both tests, i.e., the same absolute measurements from both *O*s for all *S*s and/or all conditions. Conventional use of correlation coefficient provides an estimate of relative reliability only.

For the purposes of the present experiment, computations of relative reliability estimates were considered adequate. One such estimate of reliability can be made by testing the consistency of *O*s' behavior with different *S*s. The variance attributable to interaction of *O*s with *S* was compared with the triple interaction (error term) in an *F* ratio. With the use of the variance estimates in this way, reliability is suggested when the *F* is insignificant, i.e., when *O*s are consistent with *S*s.[5] The *F* values for the seven response measures are shown in the first column of Table 4. The only *F* ratio which is significant at the .05 level (indicating lack of reliability) is the one computed for "hand movement duration."

A second determination of relative reliability may be obtained from the variance due to interaction of *O*s with conditions. The question here is whether the *O*s behave with relative consistency under the two conditions. This first-order interaction was also tested with the triple interaction as the error term. The resulting *F*'s are shown in the second column of Table 4. None of these *F*'s is significant.

The results of all these statistical tests suggest that *O*s were relatively consistent in their observations of different *S*s under the two conditions. The single exception, duration of hand movement, is noted above.

Main effects. The variable of primary interest in the study is the speech condition. In order to evaluate the effect of the two conditions on each of the responses, the variance between conditions was tested against the triple interaction.[6] The "unadjusted" *F* values in the first column of Table 5 were computed in this way. ("Adjusted" *F*'s will be explained below.)

TABLE 4

F VALUES OBTAINED FROM ANALYSES OF VARIANCE ON EACH OF THE SEVEN RESPONSE MEASURES FOR ESTIMATING "RELIABILITY"

RESPONSE CLASS	*F* FOR INTERACTION OF *O*s WITH *S*s (*df* 19, 19)	*F* FOR INTERACTION OF *O*s WITH CONDITIONS (*df* 1, 19)
Duration eye contact loss	2.11	0.05
Duration hand movement	2.68*	0.18
Duration body movement	0.91	0.14
Duration total speech	1.60	1.11
No. eye contact losses	1.19	1.09
No. hand movements	1.49	1.48
No. body movements	1.11	4.31

* $p < .05$.

Note.—*F* values whose $p < .05$ indicate lack of reliability.

The unadjusted *F* values fail to approach significance with only one of the response classes, viz., the duration of time which the listener "moves his hands." The other responses are significant at the levels indicated in the tables. From these results we get a picture of the listener during normal speech and during stuttering. He loses eye contact with the stutterer for a longer period of time than with the normal speaker. The intervals of loss of eye contact are further exaggerated by the fact that the number of times the *S* looks away is less with the stutterer than with the normal speaker. In other words, during stuttering, loss of eye contact by the listener is maintained for longer *continuous* intervals, and for a longer total duration. The amount of hand movement does not change with a change

[5] It may be argued that this statement of reliability is severely limited because it is based on failure to reject the null hypothesis. In the present instance, this statistical limitation is not serious since the main purpose of computing reliability here is to determine whether *O*'s inconsistency is significant enough to be considered in choosing an error term for testing main effects.

[6] Choice of triple interaction as the error term was guided by the fact that *O* by condition interactions failed to approach significance. This does not exclude the possibility that *S* by condition interaction may account for some of the variability between conditions. However, this interaction contains variance because of counterbalancing the order of conditions and *E*s, so that its use as an error term was considered too severe.

TABLE 5

F RATIOS BETWEEN SPEECH CONDITIONS OBTAINED FROM ANALYSES OF VARIANCE AND COVARIANCE OF THE RESPONSES

RESPONSE CLASS	UNADJUSTED *F* (*df* 1, 19)	ADJUSTED *F* (*df* 1, 18)
Duration eye contact loss	54.27***	0.46
Duration hand movement	0.09	0.41
Duration body movement	29.46**	0.02
Duration total speech	624.18***	—
No. eye contact losses	23.47***	7.24*
No. hand movements	13.73**	0.00
No. body movements	39.00***	0.00

* $p < .05$.
** $p < .01$.
*** $p < .001$.

in condition but the frequency with which the listener will initiate a hand movement decreases during stuttering. Other bodily movement is significantly depressed during stuttering both in duration and in the number of times a movement is initiated. The listener starts a "bodily movement" less frequently and for a shorter interval during stuttering. Finally, the total amount of conversation during stuttering is depressed. Whether this depression of speech is due to the *S*, the *E*, or both was not measured in this study.

The purpose of making the latter measurement, i.e., total duration of speech, was to demonstrate what differences would exist if the total speech time could be held constant. The adjusted *F*'s in Table 5 were computed by employing this measurement as the "uncontrolled variable" in a covariance design (9). The results indicate that differences between conditions remain with only one measurement, viz., number of eye contact losses. In other words, if there had been the same amount of speech in each condition, frequency of eye contact loss alone would remain significantly different between conditions.

DISCUSSION

The results of this study demonstrate that listener responses are altered significantly by stuttering. The responses studied were chosen since they constitute a part of many stutterers' complaints about their audience. More subtle and complex listener reactions are also reported by the stutterer as "unfavorable" listener behavior. Self-descriptions of these reactions indicate "impatience, amusement, embarrassment, excessive solicitude" (2, p. 297) as some of the listener responses to stuttering. Although somewhat ambiguous, it is possible that a part of the referents for some of these terms include the responses measured in this experiment. Various temporal combinations (as well as total frequency and duration) of the responses may be crucial to the report of the stutterer, and more important, to its effect upon his speech. Future experimental studies on the effect of "listener behavior" on stuttering which employ the results of this study may profitably include consideration of such a possibility.

Nevertheless, the statistical significance obtained for most of the measured responses demonstrate the probable reality of the reports of the stutterer about these particular listener reactions. This conclusion should be qualified to include the fact that there is considerable variability between *S*s. Inspection of the *SD*'s in Tables 2 and 3 leads us to believe that some extreme *S*s during "normal speech" behave quantitatively similar to other *S*s during stuttering. With this degree of overlap between conditions, some *S*s will be relatively "poor listeners" regardless of who is talking to them. Nor have the present results necessarily demonstrated that all *S*s react differently with a stutterer than with a normal speaker. Thus, indiscriminate complaining by some stutterers about "poor listeners" is not ruled out by the present results. To be accurate about a single listener, the stuttering individual must compare the same listener under two stimulus conditions, i.e., when there is relatively fluent speech and when there is stuttering. In this connection, it would be worth while to investigate how well stutterers can discriminate the two conditions by observations of the listener only.

The possibility of adaptation by the listener should not be neglected in our account of the stutterer's sensitivity to "poor listeners." It is conceivable that prolonged contact with a stutterer may reduce the differences in reactions to stuttering and normal speech. A

stutterer may continue to respond to his listener on the basis of his initial contacts in spite of subsequent changes in the behavior of the listener. The present study was not designed to study the adaptation phenomenon. More stringent controls on the spacing of interruptions, etc. would have been necessary. The reports of the *E*s are suggestive, however, of the value of such a study. They report that the *S*s appeared to "relax" and become progressively more verbal during the "stuttering conditions." The effect was not so marked during the "normal speech condition." The finding by Ainsworth (1) that the variability of breathing of the listener decreased during the study, tends to support the qualitative observations of the *E*s in the present study concerning adaptation.

A covariance analysis was made for all *S*s' responses to discover what effects would remain significant if duration of total speech were held constant. It may be argued, for example, that less verbal exchange between two persons is followed by less eye contact, less activity, etc. The covariance results indicate that only one response, frequency of eye contact loss, remains significant between conditions. This result is consistent with the possibility that decreased speech results in changes of most of the responses. However, we must not neglect the equally tenable interpretation that speech and the other measured activity covary concomitantly with changes in the independent variable. In order to experimentally control the amount of verbal behavior, it may be necessary to eliminate the difference between conditions.

In focusing on the listener, the present study has departed from most investigations in the field which are concerned primarily with the behavior of the stutterer. The utilization of the present results for further studies on the conditions under which stuttering is altered is of importance. But it may also be profitable to dwell on the behavior of the listener for a better understanding of the stuttering phenomenon. The present experiment indicates that stuttering seems to act as a behavioral depressant. In terms of learning principles, stuttering may be considered as a possible negative reinforcer to the *listener*, i.e., a noxious or punishing stimulus. Although the effects of noxious stimuli on behavior are complex, there is some reason (4, 5, 8) to believe that noxious stimuli may act as behavior depressants in some situations. This fact is particularly evident in instances where the experimenter has not intentionally provided escape or avoidance responses. In a study involving the administration of inescapable shock to rats, Brown and Jacobs (4) have hypothesized that the "freezing" of rats on the grid may actually constitute partial or complete shock avoidance. They point out that shock reduction may be effected by greater contact area of the animal with the grid. If the present experiment is an analogous situation, decrease in listener activity may be part of a noninteractive pattern which would presumably be reinforced by producing less interactive behavior on the part of the stutterer.

Evidence for such a hypothesis is provided by more than the fact that listener behavior was generally depressed during stuttering. In addition, loss of eye contact is increased during stuttering.[7] Such responses successfully terminate (avoid) at least part of the stuttering behavior (particularly visual components) performed by the *E*. A clear-cut experimental demonstration of the reinforcing properties of stuttering would involve the conditioning of a listener response in which cessation of stuttering is contingent upon the occurrence and/or maintenance of this response. Adaptation to the stuttering by the listener would also appear to be an important consideration in attempting to use stuttering as a reinforcer. Although focused on the listener, this type of study may give us a better understanding of how the stutterer alters his social interactions.

SUMMARY AND CONCLUSIONS

The purpose of this study was to investigate the effects of stuttering on the frequency and duration of (*a*) loss of eye contact, (*b*) hand movement, and (*c*) other bodily movement of the listener. Twenty college students were used as *S*s in the investigation. The experimental situation was defined for them as a waiting period *prior* to the "experiment for which they signed up." The *S*s were exposed to five min-

[7] Although not measured quantitatively, the two *E*s faking stuttering frequently reported that their sentences were completed by the listener. This is consistent with the hypothesis that stuttering is a noxious stimulus and that sentence completion is one of the instrumental acts which terminate such stimulation.

utes of normal speech in conversation with one *E* and five minutes of stuttering speech in conversation with a second *E* in this waiting period. The *E*s were presumably other *S*s who were also waiting their turns for the experiment. Observations of the *S*s' behavior were made without their knowledge through a half-silvered mirror by two *O*s.

Statistically significant *F*'s were found in five of the six response measures. The following differences in listener responses were found to occur during stuttering: (*a*) increased duration of loss of eye contact, (*b*) decreased frequency with which a change in eye contact away from the stutterer was made, (*c*) decreased frequency with which a hand movement was initiated, (*d*) decreased total duration of other bodily movement, and (*e*) decreased frequency with which a bodily movement would be initiated. Adjusted *F*'s were then computed by covariance using total speech time of *S* and *E* as the uncontrolled variable. One adjusted *F* remained significant, viz., frequency of eye contact loss.

Results of the experiment were discussed in terms of the complaints of the stutterer about his audience behavior. Also discussed were the implications of the present results in the understanding of the social interactions of the stutterer.

The results support the following conclusions:

1. Listener behavior is significantly affected by stuttering.

2. Statistical control for total amount of speech removes most of the differences in listener behavior during stuttering.

3. Stuttering seems to act as a behavioral depressant to the listener. The "avoidant" activities of the listener may be of significance in classing stuttering as a negative reinforcer (noxious stimulus) to the listener.

REFERENCES

1. AINSWORTH, S. Studies in the psychology of stuttering: XII. Empathic breathing of auditors while listening to stuttering speech. *J. Speech Disorders*, 1939, **4**, 149–156.
2. BLOODSTEIN, O. Conditions under which stuttering is reduced or absent: a review of literature. *J. Speech Hearing Disorders*, 1949, **14**, 295–302.
3. BLOODSTEIN, O. A rating-scale study of conditions under which stuttering is reduced or absent. *J. Speech Hearing Disorders*, 1950, **15**, 29–36.
4. BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
5. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57**, No. 3 (Whole No. 263).
6. FLETCHER, J. M. *The problem of stuttering*. New York: Longmans, Green, 1928.
7. GREENE, J. S. *I was a stutterer*. New York: Grafton, 1932.
8. SKINNER, B. F. *The behavior of organisms*. New York: D. Appleton-Century, 1938.
9. SNEDECOR, G. W. *Statistical methods*. (4th Ed.) Ames, Iowa: The Collegiate Press, 1946.
10. VAN RIPER, C. The influence of empathic response on frequency of stuttering. *Psychol. Monogr.*, 1937, **49**, No. 1 (whole No. 217) 244–246.

Received March 28, 1953.

COMPLEXITY OF RESPONSE AS A FACTOR IN THE VOCABULARY PERFORMANCE OF SCHIZOPHRENICS

R. HARRINGTON AND J. C. EHRMANN[1]

Veterans Administration Hospital, Marion, Indiana — *Michigan State College*

COMPLEXITY of the response required, as a factor in the vocabulary performance of abnormals, has largely been neglected in the literature. Utilizing the idea that vocabulary seemed to be the most stable aspect of intelligence, Babcock (1) devised a mental efficiency test comparing the vocabulary level with achievement on other types of tests. Yacorzynski, however, has pointed out that this apparent stability may be an artifact of the scoring systems used. He feels that definitions reflecting different kinds of thinking may receive the same quantitative score and, in relation to the fact that this score is often the same, he states, "It seems logical to assume, however, that the reason why the vocabulary of the deteriorated individual remains relatively unchanged is because easier methods are left to the organism even if the more difficult solutions are no longer available" (7, p. 265). Moran, Moran, and Blake (5), however, reject this hypothesis and suggest that most of the differences reported by Feifel were contributed by the organics and manic-depressives in his particular sample (3).

Defining a word verbally is a highly complex task involving memory, motor behavior, perception, motivation, a knowledge of sentence structure, and the ability to select and organize words in a meaningful way. Selecting a correct definition from a number of possible definitions may certainly be considered a much less complex task than offering a spontaneous verbal definition. Motivation, perception, and memory are all involved, but the selection and organization of the words into a meaningful sequence have been done by someone else. All that remains for the subject is to recognize the words and make a choice limited to these recognized possibilities.

It would appear then that the question of whether the "more difficult" solutions are no longer available to an abnormal individual is deserving of study. The present study attempted to investigate the role which the complexity of response plays in the ability of one group of abnormals (schizophrenics) to engage in abstract thinking as inferred from their performance on vocabulary tests. The hypotheses were:

1. The relative concreteness of the definitions of vocabulary items given by schizophrenics is in part a function of the complexity of the response required.

2. Schizophrenics will give significantly less abstract definitions to vocabulary words than will normals when the response required is one of verbal definition.

3. There will be no significant differences between schizophrenics and normals in the degree of abstractness of definitions of vocabulary words when the response required is the simple task of indicating a correct choice.

METHOD

Subjects. Twenty male patients and ward attendants from the United States Veterans Administration Hospital, Fort Custer, Michigan served as subjects (*Ss*). All patients had been diagnosed as Schizophrenic Reaction, Type Unclassified, and were in good contact. Each patient was equated with an attendant on the factors of: age, educational level, IQ, and vocabulary level. Table 1 presents the data on these factors.

Procedure. Two tasks were utilized. Task I called for the *Ss* to define 20 words from the Vocabulary subtest of the Wechsler-Bellevue Scale, Form I. Task II required the *Ss* to indicate the correct definitions in a series of definitions given for each word. The words were all capable of being defined as nouns and seemed to present a wide range of difficulty.

Each of the series of five definitions for each word in Task II included one each of the following types of definitions, arranged in random order (examples are given for the word "nitroglycerine"):

Type 1. A definition which would be scored as categorical according to Gerstein's examples (4): "A type of explosive."

Type 2. A definition which would be scored as functional according to Gerstein's examples: "It is used in bombs."

Type 3. A definition which would be scored as concrete according to Gerstein's examples: "It is yellow stuff."

Type 4. A definition which was in no way correct: "A large baby-buggy."

Type 5. A definition which was incorrect but might be mistaken for the correct definition: "Used to find nitrogen in the air."

[1] From Veterans Administration Hospital, Fort Custer, Michigan.

TABLE 1

MEAN SCORES ON AGE, IQ,* EDUCATIONAL AND VOCABULARY LEVELS OF THE TWO GROUPS

	ATTENDANTS	PATIENTS	DIFF.	t	p
Age (in years)	29.1	28.3	.8	.72	NS
Education (grade)	10.6	10.7	.1		
IQ	103.25	101.0	2.25	1.81	NS
Vocabulary level (standard scores)	6.5	6.8	.3		

* From Wechsler-Bellevue Intelligence Scale, Form I (6).

TABLE 2

MEAN SCORES OF THE ATTENDANT AND PATIENT GROUP ON VERBAL AND MULTIPLE-CHOICE TASKS

	ATTENDANTS	PATIENTS	DIFF.	t	p
Verbal					
Rater I	3.47	3.32	.15	2.63	.01 < .02
Rater II	3.47	3.29	.18	3.05	< .01
Multiple-choice	3.80	3.81	.01		

The purpose of including definitions of types 4 and 5 was to give some measure of whether the individuals responded in a random manner or chose incorrect definitions which were likely to be confused with correct ones. The data indicate that the scores were not materially affected by such responses.

Scoring. For Task I, the verbal definitions were rated by two independent judges according to the definitions of concrete, functional, and categorical as given by Gerstein:

> (a) *The concretistic or descriptive method.* . . . It implies memory of an object at the concrete sensory level. For example, "an apple is red."
> (b) *The functional or usage method.* . . . A word is defined by the subject recalling the use to which that object was put in the past; e.g., "an apple is something you eat."
> (c) *The categorical or conceptual method.* This represents the abstract method of definition. . . . "An apple is a fruit." is a conceptual definition whose *modus operandi* is more complex than the descriptive or functional methods" (4, p. 368).

It is important to realize that these definitions were scored on the degree of abstractness of the definitions, whether correct or not in the usual sense. Hence, a definition which would receive no credit according to Wechsler's criteria could be given an appropriate score, depending on whether it was judged to be concrete, functional, or categorical.

Each type of response received a scoring weight as follows: Don't Know—1; Concrete—2; Functional—3; Categorical—4. After scoring each definition according to the highest type of response given, an average score was obtained for each subject. In 91 per cent of the definitions given, the two judges agreed on the exact score to be assigned.

For Task II the following types of definitions gained scores as indicated: No response—1; Incorrect response—1; Incorrect but capable of being confused with the correct definition—1; Concrete—2; Functional—3; Categorical—4.

RESULTS

It can be seen from Table 2 that the patient group was significantly less abstract in its verbal definitions than was the attendant group. This finding is in accord with what would be expected on the basis of the work of other investigators and supports hypothesis 2 of this study.

Table 2 also indicates that there was no significant difference between the groups on the multiple-choice task, and thus hypothesis 3 of the present study is supported. The results of this study fail to show the differences obtained by Chodorkoff and Mussen (2).

These results lend support to the Yacorzynski hypothesis and cast doubt upon the explanation given by Moran, Moran, and Blake (5) in regard to Feifel's results. When analyzed on the present three levels, the schizophrenics did give significantly fewer abstract definitions than equated normals, and it is entirely possible that the same phenomenon occurred in Feifel's sample.

Both groups obtained a higher score on the multiple-choice task than on the verbal task, this difference being significant for both groups at beyond the .01 level. This discrepancy between the scores on the verbal and multiple-choice tasks was significantly greater, at the .02 level, for the patient group than for the attendant group, when the smallest difference between the means (Rater I) is considered.

DISCUSSION

These results suggest that complexity of the response is an important factor in the vocabulary performance of schizophrenics, and the differential findings of this investigation raise some interesting questions regarding the cognitive functions in schizophrenia. Whether they can be explained as an attempt by the schizophrenic to simplify, and thus maintain, a semblance of order in the world, or whether they represent a definite impairment in the ability to give a complex, highly integrative response is a problem for future investigation.

SUMMARY

Equated pairs of schizophrenics and attendants were tested on two tasks differing in the

complexity of the response required. The performance of the schizophrenic group was significantly poorer than that of the attendants where the more complex response was involved. Where the less complex response was involved, the groups did not differ.

Both groups did significantly better on the less complex task but this discrepancy was significantly greater for the schizophrenic group.

The results are interpreted as supporting the hypotheses of the study. They support the Yacorzynski hypothesis and cast doubt on Moran, Moran, and Blake's recent explanation of Feifel's results.

REFERENCES

1. BABCOCK, H. An experiment in the measurement of mental deterioration. *Arch. Psychol.*, 1930, No. 117.
2. CHODORKOFF, B., & MUSSEN, P. Qualitative aspects of the vocabulary responses of normals and schizophrenics. *J. consult. Psychol.*, 1952, **16**, 43–48.
3. FEIFEL, H. Qualitative differences in the vocabulary responses of normals and abnormals. *Genet. Psychol. Monogr.*, 1949, **39**, 151–204.
4. GERSTEIN, REVA A. A suggested method of analyzing and extending the use of Bellevue-Wechsler vocabulary responses. *J. consult. Psychol.*, 1949, **13**, 366–374.
5. MORAN, L. J., MORAN, FRANCES A., & BLAKE, R. R. An investigation of the vocabulary performance of schizophrenics: II. Conceptual level of definitions. *J. genet. Psychol.*, 1952, **80**, 107–132.
6. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
7. YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock deterioration test. *Psychol. Rev.*, 1941, **48**, 261–267.

Received April 6, 1953.

AUTHORITARIANISM AND LEADERSHIP CHOICE IN A MILITARY SETTING[1,2]

E. P. HOLLANDER[3]

Carnegie Institute of Technology

SINCE the publication of *The Authoritarian Personality* (1) a few short years ago, a spate of studies has been reported in which scores derived from the F scale or one of its modifications have been related to a variety of cognitive, behavioral, or social criteria. Paralleling this development, the influence of situational determinants of leadership has received heightened emphasis in the literature (e.g., 8, 11, 18). A number of recent reports, in particular, have sought to bring about an articulation of these two areas by relating authoritarianism[4] to leadership emergence and followership (2, 4, 7, 16, 22). Among other evidence contained in these reports, Sanford (16) has provided a link between authoritarianism and orientation to leadership, and Christie (4) has established a relationship between acceptance within military trainee groups in an Army basic training unit and shifts in authoritarianism scores accompanying training. The findings of these studies suggest that an individual's choice of a leader is related to his authoritarianism and that, within a military context, individuals who show an increase in their authoritarianism scores during training tend to be more readily assimilated into the organizational structure.

PROBLEM

Within military groups, the problem centering about authoritarianism and leadership choice by peers poses at least three fundamental questions: First, is the military leadership role perceived to be authoritarian within a given military group? If so, are the more authoritarian members of the group more likely to be chosen as "military leaders" by their potential followers? Finally, is there a definable relationship between one's own authoritarianism score and the authoritarianism score of the individual who is one's leadership choice?

An answer to the first question has been tentatively provided by an exploratory study recently completed by the author among Naval Aviation Cadets. The hypothesis tested was that the military leadership role would be perceived by these subjects to be authoritarian, within a framework provided by the F scale. It was found that, when the cadets were given two consecutive administrations of the F scale, the first under a conventional set and the second with instructions to respond to the scale *as though* it were "a test of military leadership potential" on which they wished to make a "good score," a significant upward shift in F scores was obtained from the first to the second administration. The hypothesis was upheld, and it was concluded that F-scale attitudes were likely to be ascribed to the military leadership role by the subjects utilized. The two remaining questions were then investigated by the study to be reported here. The hypotheses considered were as follows: (*a*) that there is a significant positive relationship between scores achieved on the F scale and choice by one's peers for a military leadership position; and (*b*) that there is a significant difference, with respect to F-scale scores, between the leadership nominees of those

[1] Opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

[2] This report is based upon a dissertation submitted to Columbia University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to acknowledge his appreciation to Drs. Goodwin Watson, Robert L. Thorndike, Edward J. Shoben, Otto Klineberg, Kenneth Herrold, and Samuel Flowerman of that institution, and to a multitude of colleagues at the Naval School of Aviation Medicine, the Bureau of Medicine and Surgery, and the Office of Naval Research for their valuable assistance and encouragement in the implementation and fulfillment of this study.

A paper drawn from the substance of this report was presented before Division 8 at the 1953 meeting of the American Psychological Association.

[3] The research covered herein was completed by the author while at the Aviation Psychology Laboratory of the U. S. Naval School of Aviation Medicine, Pensacola, Florida.

[4] The term "authoritarianism" as it is used here shall generally be restricted in meaning to an acceptance of authoritarian attitudes embodied in the F scale or one of its modifications.

"high" on authoritarianism and those "low" on authoritarianism.

METHOD

Subjects. Nine consecutively formed sections of cadets at the Naval School of Pre-Flight, a total *N* of 268, were drawn as a sample. The age range for the group was 19 to 26; its mean age was 22.1 years, with slightly more than 80 per cent of the cases falling within a range of 21 through 24. The educational level of the subjects (*S*s) ranged from high school graduates to one *S* with six years of college training; the mean was 2.8 years of college, with approximately 86 per cent of the group having had two or more full years of college. About half of the group had had some previous enlisted service. Geographically, the process of selection operated so as to represent the major areas of this country in the study sample.

Procedure. At the end of three months of training, the *S*s were asked to fill out a leadership nomination form for their section on which they indicated three cadets whom they considered to be *best* qualified for the hypothetical position of "student commander," and three whom they considered to be *least* qualified. This form closely approximates those sociometric nomination forms used in previous studies in similar settings (e.g., 14, 20, 21). By applying appropriate plus and minus weights of 3, 2, 1 for each nomination an individual received, a leadership score was derived. At about the same time in the test administration schedule, *S*s were requested to complete the F scale. The leadership scores and F scores were then transformed to standard scores based upon the individual section.[5] In each case, 30 was set as the mean and 10 as the *SD*.

RESULTS

The major hypothesis was tested by two analyses. The first of these, a chi-square analysis making use of the total sample distributed into categories of "high," "medium," and "low" for the two major variables, is presented in Table 1. As will be noted, the chi square was significant beyond the .05 level, with the obtained frequencies indicating a rejection of the hypothesis. In addition to this, a correlational analysis was completed between authoritarianism and leadership scores for the total sample. The correlation of −.23 secured between these variables was significant at the .01 level.[6]

TABLE 1

AUTHORITARIANISM SCORES AND LEADERSHIP STANDING BY CRITERION GROUPS†

AUTHORITARIANISM	LEADERSHIP*			TOTALS
	HIGH	MEDIUM	LOW	
High	13 (21)**	34 (31)	27 (22)	74
Medium	30 (32)	53 (47)	31 (35)	114
Low	32 (22)	25 (34)	23 (24)	80
Totals	75	112	81	268

† *High* indicates a score of .6σ or more above the section mean; *Medium* indicates a score between .5σ above and .5σ below the section mean; *Low* indicates a score of .6σ or more below the section mean.

* Chi square = 12.77; $df = 4$; $p < .05$.

** The numbers in parentheses indicate the expected frequency for each cell.

In a supplementary series of analyses, it was found that scores achieved by *S*s on the college-level ACE test correlated negatively with the authoritarianism scores, $r = -.21$, and positively with leadership, $r = +.30$. Accordingly, a partial r was calculated between the two major study variables with intelligence held constant; $r_{AL \cdot I}$ was found to be −.18, a low but still significant relationship. No other variables studied, i.e., age, educational level attained, or previous military experience, were found to be related significantly to both authoritarianism and leadership. Officer-like qualities (OLQ) scores assigned by superior officers in charge of the cadets were found, however, to be correlated with the leadership scores at a level of +.55. On the other hand, the authoritarianism scores correlated with OLQ at a virtually zero level, $r = -.06$. Table 2 presents a matrix of the relevant intercorrelations.

The second hypothesis was tested by a further analysis summarized in Table 3. Those "high" on authoritarianism were compared with those "low" on authoritarianism with regard to their respective leadership nominees' scores on the authoritarianism variable. The results revealed that the authoritarianism scores of those *S*s nominated "highest" on leadership, by *S*s who were respectively "high" and "low" on authoritarianism, did not significantly differ from

[5] The distribution of raw F-scale total scores was unimodal and approximately symmetrical. The mean equalled 114.01 with an *SD* of 20.92. There is no reason to suppose that this sample differed significantly from the California samples on which the F scale was normed.

[6] Since some degree of curvilinearity was observed in the correlation plot, an eta was computed and found to be .29. The chi-square test of linearity was applied, and a value which did not support the hypothesis of nonlinearity was yielded.

TABLE 2

INTERCORRELATIONS AMONG AUTHORITARIANISM SCORE, ACE TEST SCORE, OFFICER-LIKE QUALITIES SCORE, AND LEADERSHIP CRITERION FOR TOTAL STUDY SAMPLE

($N = 268$; $\sigma_r = .06$)

VARIABLES	ACE	OLQ	L. CRIT.
Authoritarianism	−.21**	−.06	−.23**
ACE		+.17**	+.30**
OLQ			+.55**

** $p < .01$.

TABLE 3

MEAN AUTHORITARIANISM SCORES* OF CADETS NOMINATED HIGHEST OR LOWEST ON LEADERSHIP BY CADETS WHO WERE HIGH OR LOW ON AUTHORITARIANISM

NOMINATORS	NOMINEES		
	NOMINATED "HIGHEST" ON LEADERSHIP	NOMINATED "LOWEST" ON LEADERSHIP	
High score on authoritarianism	$\bar{X} = 25.30$ $\sigma^2 = 80.06$ $N = 74$	$\bar{X} = 33.74$ $\sigma^2 = 136.97$ $N = 72$	$CR = 4.87$ $p < .001$
Low score on authoritarianism	$\bar{X} = 27.29$ $\sigma^2 = 60.60$ $N = 80$	$\bar{X} = 33.17$ $\sigma^2 = 129.55$ $N = 78$	$CR = 3.79$ $p < .001$
	$CR = 1.38$ $p > .10$	$CR = .30$ $p > .10$	

* The scores reported are based upon a standard score distribution with a mean of 30 and an *SD* of 10.

one another; this held true for the "lowest" nominees as well. It would appear, therefore, that *S*s nominated as "highest" on leadership were significantly lower on authoritarianism, irrespective of the standing on authoritarianism of those making the nominations.

DISCUSSION

The findings obtained indicate a rejection of the two hypotheses with the resultant inference that authoritarians are less acceptable, as leaders, to their fellow cadets, than are nonauthoritarians. A fundamental paradox seems to be posed here. The military institution does present an essentially authoritarian structure after all. Moreover, it may be granted that military officer indoctrinees are exposed to an authoritarian conception of military leadership,[7] and, in keeping with the evidence presented above, that this cadet population tends to ascribe authoritarian attitudes to the military leadership role. Yet, in a free choice situation such as this, where a military leadership role is to be filled, the more authoritarian individuals are not popular choices among the trainees.[8] Indeed, precisely the reverse trend is observed. In considering this finding, there are two assumptions involved which properly might be scrutinized: (*a*) that *S*s would act upon the *traditional* role prescriptions which others like themselves appear to recognize as bound up with the military leadership role; (*b*) that the acceptance of authoritarian ideology, as embodied in the F scale, necessarily implies the manifestation of some form of authoritarian behavior.

Taking account, first, of the discrepancy between the recognition of certain role prescriptions and the actual selections for role fulfillment, one might contend that the conception of an authoritarian military leader may be anchored in an incomplete and somewhat distorted stereotype. Operationally, this stereotype might be rejected under the impact of the intimate human interactions which serve to determine leadership choice, even within a military context. A case might readily be made in support of the further contention that the individual chosen as a leader by his peers, in a military setting where rigid lines of authority are drawn, may be one who is less authoritarian in his interactions with them.[9] This relationship might be attributable to loyalties within the peer group which resist dominance behavior of the sort of which the authoritarian is presumptively capable. Thus,

[7] Cf. M. Brewster Smith's description of the impact of OCS in *The American Soldier* (19, p. 389).

[8] While this finding is apparently in contradiction to Christie's (4), it is of special interest to note that his most accepted group had initially been *below* the group mean on authoritarianism.

[9] Although this would appear to run contrary to the work of Lewin and his associates (12), where internal aggression was found to characterize groups of youngsters under authoritarian rule, there may be several obvious reasons which would serve to explain this contradiction: First, cadet groups function with a high level of goal orientation; second, numerous opportunities are afforded for the release of aggression through competition with other groups; finally, there remains an open question as to the comparability of child and adult groups.

the authoritarian may elicit a negative response from his potential followers because of his emulation of an authoritarian conception of the military leadership role.

There yet remains a question as to the reasons for the conflict between this study and that by Carter, Haythorn, *et al.* (3) which found that participants who secured leadership status in leaderless group discussions displayed authoritarian behavior. Notice, first, that in the latter study a behavioral criterion was applied and, second, that the investigators studied discussion groups of an unstable, temporary nature. The current study simply *infers* authoritarian behavior and, in contrast, makes use of persisting "real" groups of some months' establishment. It might also be noted that Bass, McGehee, *et al.* (2) report a significant negative, but somewhat curvilinear, relationship between F-scale scores and leadership status in leaderless group discussions. The discrepancy between this finding and the one reported by Carter, Haythorn, *et al.* leads appropriately to a consideration of the second assumption noted in the foregoing, that is, the correspondence between authoritarian behavior in some form and the subscription to F-scale attitudes.

Without implying any disrespect for the monumental achievement represented by *The Authoritarian Personality*, one might wish nonetheless to further examine the F scale and the meaning of an F-scale score. At the outset, fairness demands that consideration be given the fact that the scale's constructors were concerned with tapping the personality syndrome which would be particularly receptive to antidemocratic ideology (1, p. 4). In this regard, they have established that the scale is related to clinical evidence of rigidity, stereotypy of thinking, moral conservatism, and the like. Moreover, the scale has been shown to correlate respectably with scales of anti-Semitism, ethnocentrism, and political-economic conservatism. Yet, withal, there exists no body of evidence which directly indicates that an acceptance of this ideology is necessarily indicative of some qualitative behavior, i.e., a traditionally conceived "authoritarian behavior." This is not to deny that behavioral differences can be inferred (as they usually are) from empirical evidence. Several notable studies, among them those by Rohde (15), Bass, McGehee, *et al.* (2), and Scodel and Mussen (17), support the notion that perceptible differences do exist between the behavior of "high" and "low" scorers on the F scale. Even where the influence of intelligence is held constant, as is the case in the current study, apparent differences in the behavior of "highs" and "lows" persist. This kind of evidence, while not conclusive, cannot be dismissed lightly. But it still gains us little in our quest for the precise behavioral qualities of the extreme groups on a continuum of F-scale scores—which means, in effect, that we cannot necessarily impute authoritarian behavior to the person who accepts F-scale attitudes.

What may we say, then, about the individual who scores high on the F scale? In the first place, it is apparent that these "authoritarians" are defined as such on the basis of their having agreed with some grouping of F items. Furthermore, as they stand, these items appear to be a potpourri of discredited concepts running the gamut from hokum ("some day it will probably be shown that astrology can explain a lot of things") to sanctimonious self-indulgence ("no sane, normal, decent person could ever think of hurting a close friend or relative"). When this state of affairs is taken into account, it seems reasonable to consider that the "high" scorer on the F scale may not be so much an authoritarian, behaviorally, as he may be an individual with probable inadequacies in the sphere of social intelligence or social perception. This conception does no great disservice to the work of the California group, and it partially explains contradictions of the sort alluded to above in connection with the study by Carter, Haythorn, *et al.* (3) on the one hand, and Bass, McGehee, *et al.* (2) on the other. More pointedly, too, it brings into focus a likely explanation for the findings obtained in the current study.

A variety of investigations (cf. 11, 18) have suggested that responsiveness to the needs of others, a social facility perhaps, may constitute a relatively stable component of leadership. Also, studies completed in military units have indicated that some factor similar to "permissiveness" plays a role in determining a leader's acceptability to his followers. In

this regard, particular attention is directed to a study reported in *The American Soldier* (19, p. 385) and to an Air Force study carried on during World War II (5, p. 34). Both the enlisted and officer combat veterans questioned in these studies indicated a conception of the military leader's role which was heavily weighted in the permissive direction. Of special note was the expression of a desire for greater responsiveness to their needs, opinions, and very beings by the men appointed over them.

If it is true that the man who subscribes to the bulk of the F-scale items is correspondingly inadequate in his social interactions, this would imply that he is ill-equipped to accept the needs of other individuals or to deal with them in a socially effective fashion.[10] Even within the context of an essentially authoritarian institution, therefore, the "F-scale authoritarian" may be expected to draw only a minimal following for a position of leadership. Viewing the findings within this frame of reference probably serves best to explain the fact that whether nominators were "high" or "low" on the F scale, they tended to favor "low" scorers as military leaders.

In conclusion, then, we have seen that acceptance of authoritarian attitudes may make an individual less desirable as a military leader from the standpoint of those who may be placed in a position of followership—in this case, officer indoctrinees. Whether such free choice of a leader renders that leader effective in institutional terms remains an open question to be studied further, although there already exists some affirmative evidence (e.g., 20, p. 157; 21, p. 291).[11] In this vein, additional study appears indicated as well with respect to the relationship between the acceptance of F-scale attitudes and the manifestation of authoritarian behavior. Clearly, we are confronted with the fact that there are a variety of meanings to the descriptive term "authoritarian," i.e., an ideology, a behavior, and an institutional form. At this juncture, the reality of the matter appears to be that these meanings cannot be interchanged without attendant confusion in our research efforts.

It should be reiterated that the relationships found in this investigation are of a relatively low magnitude and that their utility for prediction is virtually nil. On the positive side, however, there remains a residual of evidence which indicates that there are personal qualities relevant to social interaction, measurable to some extent by the F scale, which bear upon an individual's acceptability as a military leader to potential followers.

SUMMARY

A study was undertaken to determine the relationship between F-scale scores and leadership status among officer indoctrinees. The major hypothesis was that there would be a significant positive relationship between scores on a measure of authoritarianism and leadership acceptance by peers in a military setting.

Nine consecutive sections of cadets at the Naval School of Pre-Flight, an *N* of 268 cases, were utilized as *S*s. At the end of the third month of training, *S*s filled out a leadership nomination form for their section on which they named three cadets *best* qualified for the hypothetical position of "student commander" and three who were *least* qualified. At about the same time, cadets completed the F scale. Leadership nomination scores and authoritarianism scores were then derived for all cadets.

Correlational and chi-square analyses tested the hypothesis and indicated a significant negative relationship between authoritarianism and leadership. The correlation of −.23

[10] M. B. Jones, of the Naval School of Aviation Medicine, has recently completed a study with two independent samples of cadets in which he correlated the Taylor Manifest Anxiety Scale of the MMPI with F-scale scores using a Kendall rank correlation. The Pearson estimates for these groups were +.21 (N = 166) and +.22 (N = 245) respectively; in both cases the r's were significant beyond the .01 level. It would seem, therefore, that "high" scorers on the F scale are inclined to have greater anxiety, which undoubtedly bears on the adequacy of their social interaction. This, in turn, may be seen to accord well with the points raised relative to the leadership acceptability of such individuals.

[11] A recent study by the author (10) indicates that peer nominations on leadership, taken at the preflight level, significantly predict the pass-fail criterion in Naval Air Training. The validity coefficient for this variable was equalled only by that for final preflight average; other predictor variables from this level of training, including superiors' ratings on military leadership qualities, showed no comparable predictability.

between these variables was significant beyond the .01 level. A supplementary series of analyses focused upon the relationship of a number of other variables to the two main study variables. Intelligence, as measured by the ACE test, was found to be related significantly to both authoritarianism, $r = -.21$, and peer nominations for leadership, $r = +.30$. The significance of the negative relationship between authoritarianism and leadership persisted, however, even when the effect of intelligence was partialed. In addition to the major finding, it was found that the leadership nominees of those cadets "high" or "low" on authoritarianism did not differ significantly with respect to their authoritarianism scores. In both cases, those nominated "highest" on leadership had a significantly lower score on authoritarianism than those nominated "lowest" on leadership.

A number of tentative suggestions was offered to explain the relationships obtained. It was concluded that a high score on the F scale might be indicative of a lack of social intelligence or social perception. Accordingly, it was suggested that individuals who are "authoritarians" in the F-scale sense are unable to deal effectively with the needs of others and therefore tend to be rejected as leaders by potential followers. Further study aimed at pin-pointing the multiple usages of the term "authoritarian" was indicated as necessary.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BASS, B. M., MCGEHEE, C. R., HAWKINS, W. C., YOUNG, P. C., & GEBEL, A. S. Personality variables related to leaderless group discussion behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 120–128.
3. CARTER, L. F., HAYTHORN, W., MEIROWITZ, B., & LANZETTA, J. The relation of categorizations and ratings in the observation of group behavior. *Hum. Relat.*, 1951, **4**, 239–253.
4. CHRISTIE, R. Changes in authoritarianism as related to situational factors. *Amer. Psychologist*, 1952, **7**, 307. (Abstract)
5. CRANNELL, C. W., & MOLLENKOPF, W. G. Combat leadership. Chapter III in *Army Air Forces Aviation Psychology Program Research Reports.* Washington: U. S. Government Printing Office, 1947. (Rep. No. 14)
6. GIBB, C. A. The principles and traits of leadership. *J. abnorm. soc. Psychol.*, 1947, **42**, 267–284.
7. HAVRON, D., FAY, R., & GOODACRE, D. *The effectiveness of small military units.* Washington: Institute for Research in Human Relations, 1951.
8. HEMPHILL, J. K. *Situational factors in leadership. Ohio St. Univer. Stud., Bur. Educ. Res. Monogr.*, 1949, No. 32.
9. HOLLANDER, E. P. An investigation of the relationship between authoritarian attitudes and leadership selection in a military setting. Unpublished doctor's dissertation, Columbia Univer., 1952.
10. HOLLANDER, E. P. Peer nominations on leadership as a predictor of the pass-fail criterion in Naval Air Training. *J. appl. Psychol.*, 1954, **38**, 150–153.
11. JENKINS, W. O. A review of leadership studies with particular reference to military problems. *Psychol. Bull.*, 1947, **44**, 54–79.
12. LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally created "social climates." *J. soc. Psychol.*, 1939, **10**, 271–299.
13. MASLOW, A. H. The authoritarian character structure. *J. soc. Psychol.*, 1943, **18**, 401–411.
14. MCCLURE, G. E., TUPES, E. C., & DAILEY, J. T. Research on criteria of officer effectiveness. *USAF Hum. Resour. Res. Cent., Res. Bull.*, 1951, No. 51–8.
15. ROHDE, K. J. The relation of authoritarianism of the air crew member to his acceptance by the airplane commander. *Amer. Psychologist*, 1951, **6**, 323. (Abstract)
16. SANFORD, F. H. *Authoritarianism and leadership.* Philadelphia: Institute for Research in Human Relations, 1950.
17. SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181–184.
18. STOGDILL, R. M. Personal factors associated with leadership: a survey of the literature. *J. Psychol.*, 1948, **25**, 35–71.
19. STOUFFER, S. A., SUCHMAN, E. A., DEVINNEY, L. C., STAR, S. A., & WILLIAMS, R. M. *Adjustment during army life.* Vol. 1. *The American soldier.* Princeton: Princeton Univer. Press, 1949.
20. WHERRY, R. J., & FRYER, D. H. Buddy ratings: popularity contest or leadership criterion? *Personnel Psychol.*, 1949, **2**, 147–159.
21. WILLIAMS, S. B., & LEAVITT, H. J. Group opinion as a predictor of military leadership. *J. consult. Psychol.*, 1947, **11**, 283–291.
22. *Leadership identification and acceptance.* Philadelphia: Institute for Research in Human Relations, February, 1952.

Received April 1, 1953.

TEST ANXIETY AND RORSCHACH PERFORMANCE

FRANCIS N. COX
New England University College, Australia
AND SEYMOUR B. SARASON
Yale University

RECENT studies (8, 9, 10, 11) have been concerned with the analysis of drives and responses elicited in the test situation and with the experimental evidence supporting hypotheses concerning relationships between level of anxiety and quality of performance under different conditions of test administration. In light of these studies it now seems possible to attempt to extend the theoretical analysis to performance in other test situations, both to determine the generality of the hypotheses and to test the validity of certain clinical tests of anxiety. Projective techniques afford an obvious field for such experimentation, and this study is an investigation of the effects of the above variables upon Rorschach performance.

Since Benton (2) commented upon the lack of adequate experimental investigations of Rorschach anxiety indices, Eichler (4) has reported results of a carefully controlled study. His procedure was to compare the performance of two matched groups of college students on fifteen alleged Rorschach anxiety signs before and after exposure to stress conditions. " 'Stress' consisted of the administration of electric shock of increasing intensity, the implication by the experimenter of forthcoming application of stronger shock, and the use of certain equipment for suggestive effects" (4, p. 354). These conditions had a significant effect on four indices, and three others yielded suggestive results.

In the present study the performances of questionnaire-selected high and low anxious subjects (*S*s) were compared under two different conditions; half of each group was given standard instructions, and the other half was told their intelligence was being tested. Those of Eichler's anxiety signs which came out significantly or nearly significantly were used. Other predictions were derived from the assumptions to be presented below. It should be emphasized that one of our predictions is the direct opposite of one of Eichler's—namely, that relative to low anxious *S*s, high anxious individuals will give fewer surface shading (*Fc*, *cF*, and *c*) responses.[1]

THEORETICAL CONSIDERATIONS AND HYPOTHESES

The following is in part a summary from a more detailed presentation in a previous publication (8).

Assumption 1. Anxiety is assumed to be a learned drive with the characteristics of a strong stimulus. When anxiety has been learned as a response to situations involving intellectual achievement, two incompatible responses are likely to be evoked:

a. Responses which are *not* task-relevant (designated as R_A). These responses are assumed to be associated with subjective, personalized, self-centered feelings.

b. Task-relevant responses which reduce anxiety by leading to completion of the task. These responses (designated as R_{AT}) are held to be functionally equivalent to other task-relevant responses (R_T) evoked by cues in the test situation.

Assumption 2. It has also been assumed that there is a positive relationship between level of anxiety and frequency of evocation of responses of the order R_A.

Assumption 3. The stronger the drive the

[1] It will be recalled that Eichler (4) reported an *increase* in the frequency of weighted shading responses under stress conditions. Our assumption 1 that high anxiety is associated with subjective, personalized, self-centered responses implies that high anxious individuals will be so preoccupied with themselves in giving a response that they will tend to pay less attention to the stimulus material than will less self-centered, low anxious individuals. Since responding to *surface* shading seems likely to require paying considerable attention to the stimulus material, we predicted that low anxious individuals would yield proportionally *more Fc*, *cF*, and *c* responses than high anxious *S*s. Like Eichler, however, we predicted the opposite for responsiveness to *diffuse* shading (*FK* and *K*).

stronger must be the defense against the drive when the need for defense is evoked. The stronger the drive the more variable will be the inhibiting effects of the defensive reaction.

The following hypotheses are relevant to the present study:

1. Individuals with a strong anxiety drive (HA *Ss*) will yield *higher* scores than *Ss* with a low anxiety drive (LA *Ss*) on Rorschach indices which can be presumed to measure subjective response tendencies, because of evocation of more R_A responses.

2. HA *Ss* will yield *more variable* scores than LA *Ss* on Rorschach indices which can be presumed to measure response readiness, both because of evocation of more R_A responses *and* of defenses against making such responses.

3. The above differences between HA and LA *Ss* will be gradually reduced as the test progresses because of differential reinforcement of R_{AT} responses and a concomitant extinction both of R_A responses and of defenses against making such responses. It should be stressed that it is *not* predicted that the indices will become identical (and certainly not reversed), since it seems unlikely that R_{AT} responses will be as strongly reinforced in this situation as in the Kohs Block Design test (8)—objective success and observable improvement being absent from the Rorschach. It is predicted, however, that other reinforcements[2] will result in a significant increase in the frequency of R_{AT} responses and a corresponding decrease in the incidence of R_A responses among HA *Ss*. Consequently, as the test progresses, there should be a decrease between HA and LA *Ss* in the magnitude and variability of anxiety indices.

4. When a stimulus situation contains elements which specifically arouse achievement anxiety, the increase in anxiety drive *and* in defense against this drive should lead to a greater disparity between HA and LA *Ss* than when anxiety-arousing cues are minimized.

[2] Other possible sources of reinforcement which could reduce anxiety might be responses by the subject to subtle cues from the experimenter which seem to imply acceptance of his performance, or anticipated social approval because of his ability to produce some responses.

METHOD

Section I

The test anxiety questionnaire described in previous articles (11, 12) was administered to 366 students in an introductory psychology course at Yale College, and *Ss* with the highest and lowest anxiety scores were selected. This sample consisted of 40 *Ss*: 20 who scored in the upper seventeen percentiles (designated as HA), and 20 in the lower fifteen percentiles (LA) of the distribution. Each of these two groups was randomly divided into two subgroups, one of which received "ego-involving," the other standard Rorschach instructions. Approximately two months after the questionnaire was administered, these students were asked to participate in a psychological experiment. Apparently none of the *Ss* was aware of any connection between the questionnaire and the experiment.

Section II

The experimenter (*E*) had no knowledge of the anxiety classifications, merely knowing the type of instructions he had to administer to each *S*. The instructions were:

a. Standard: After spending five minutes establishing rapport in a predetermined way, *E* gave each of this group of *Ss* the following instructions: "I'll tell you what we are going to do. You know you can drop ink on a sheet of paper, fold it, squeeze it, and when you open it up get a series of differently shaped things in such inkblots—some see this, some that. There are no 'right' or 'wrong' answers because we all see different things. Well, I have got some cards made in exactly the same way, and I am going to give them to you and ask you to tell me what different things you see in them. Remember, there are no 'right' or 'wrong' answers." In the inquiry *S* was asked to trace each response after responding to the question: "What was it in the blot that made you think of . . .?" (*Verbatim* reading of response.)

b. "Ego-involved": Under these conditions there was no "rapport establishment" period, and *E* promptly gave the following instructions in a curt and incisive voice: "I want you to do a new kind of test, one which is designed to measure the range of your imagination, the quality of your intelligence. In essence, the aim of this experiment is to see how skillful you are in handling strange and novel material." The remainder of the instructions were as in the standard group except for the inclusion of the statement, "what people see depends upon their imagination, their creative capacity."

The *E* presented the first four cards making no further comments, and recorded the same data as under standard conditions. After the fourth card, however, he remarked: "Remember, this is a test of your imagination, of how creative you are. It is just because the test is strange to you that you are in a position to use your intelligence." No further comments were made until after the seventh card when the above statements were repeated. From there on, the procedure was identical with that used under standard conditions.

Section III

When all tests had been administered, *E* scored the records—still being unaware of the anxiety classifications. Scoring was after Klopfer and Kelley (7), with these modifications:

a. Responses in which no form was intended were given only the label of the appropriate determinant (*F*, *CF*, *K*, etc.). Responses so treated were rocks, clouds, fire, or sticks.

b. All other responses were given a plus or minus form level rating on the basis of Hertz's (6) norms.[3]

c. Popularity ratings were made from a combination of Klopfer and Kelley's (7) and Beck's (1) lists of *P* responses.

d. If more than one determinant was scored for a particular response, then a predetermined order was adhered to throughout: human or animal movement, color, surface or diffuse shading, and inanimate movement.

Finally, 63 responses were selected for an independent reliability check in such a way that one response came from each record and, on the average, six from each card.

RESULTS

The reliability check yielded complete agreement for scoring of main determinants in 59 of the 63 responses (93.6% agreement). In view of this high degree of compatibility and because of the necessarily arbitrary character of any system of differential weighting of secondary determinants, it was decided to restrict analysis to main determinants.[4]

After homogeneity of variance was established, the *t* test was used to determine the significance of difference in *mean* numbers of responses. Since, as is well known, the frequency of determinants is positively correlated with the number of responses, *all* other results were analyzed in the following way. First, each *S*'s score on every determinant was expressed as a percentage of his total number of responses. However, when these individual results were combined, it was found that the distributions for most indices were markedly skewed and that very few individual scores reached 50 per cent, so it was necessary to transform these individual percentages by means of the angular transformation table (5, pp. 13 and 56; 13, p. 316). To determine the extent to which these transformations had normalized the scores, the cumulative percentage distributions of each of the transformed variables were plotted upon normal-probability paper. Since the results closely approximated straight lines, it was possible to apply *t* tests to calculate the significance of the mean differences between high and low anxious *S*s on these indices. One-tailed distributions were used only when the mean difference was in the predicted direction.

The following predictions arise out of the assumptions and hypotheses stated previously, and from Eichler's (4) findings:

Prediction 1. From Eichler's data it is predicted that, under standard conditions, the high anxiety group (HA) will yield more card rejections, fewer total responses, fewer whole responses, fewer popular responses, and fewer weighted color responses than the low anxiety group (LA).

Prediction 2 (*Hypothesis 1*). From assumptions 1 and 2, it is predicted that, under standard conditions, the HA group will yield more subjective, personalized, self-centered responses than the LA group. More specifically, relative to the LA *S*s, HA individuals will yield more irrelevant verbalizations (defined operationally in terms of extraneous remarks and qualifying remarks said by the subject during the actual test); more movement (*FM*, *M*, and *m*) responses; more intended vague form responses; more poor form responses; more personal, bodily responses (sexual and anatomical content); more diffuse shading responses (*FK* and *K*); more form secondary responses (*K*, *C'*, *C'F*, *c*, *cF*, *CF*, and *C*); fewer form primary responses (*FK*, *FC'*, *Fc*, and *FC*); more repetitive, perseverative responses (defined operationally as a response given more than twice in one record); and finally, fewer surface shading responses (*Fc*, *cF*, and *c*). The logic of the derivation of the surface shading prediction has already been discussed (see footnote

[3] The *E* made his own evaluation of form level before referring to Hertz's norms, and his ratings agreed with hers in all cases where she had listed the response. Thirteen responses not tabulated in her charts were rated independently by the writers, and 11 of their assessments were identical.

[4] The significance of the results seems to justify this procedure. Further, the frequency of secondary *FK* and *K* responses was computed, and the significance of the difference between HA and LA *S*s was virtually unchanged ($p = .40$).

1), and the other expectations followed from this self vs. card orientation.[5]

Prediction 3 (*Hypothesis 2*). From assumptions 1, 2, and 3, it is predicted that, under standard conditions, HA *S*s will display *more variability* in their records than will LA *S*s. More specifically, relative to LA *S*s, HA individuals will yield more extreme results on the following indices: total number of responses; number of locations used per card; reaction times for each response; total time for each card; and finally, in the way they respond to the stimulus material (defined operationally in terms of frequency of movement of cards). All these predictions involved the assumption that these indices are likely to be reliable indicators of "response readiness" (see Hypothesis 2).

The validity of these predictions (and hence of these Rorschach indices) was tested by comparing the results of HA and LA *S*s under standard conditions. However, as will be shown later, similar results were obtained under "ego-involving" conditions. Since type of instructions did *not* have a significant effect upon performance, it was decided to combine the data from standard and "ego-involving" conditions, and these results are presented in Table 1 where the horizontal lines divide the three groups of predictions. It must be emphasized that *all* results are in the same direction when analysis is restricted to comparison of *S*s who received standard or "ego-involving" instructions.

The results relevant to Predictions 1, 2, and 3 were:

Prediction 1. Three of Eichler's predictions (card rejections, weighted color, and total number of responses) are not borne out. The inverse relationships between level of anxiety and frequency of whole and popular responses confirm Eichler's findings.

Prediction 2 is strongly supported—all ten results are in the expected direction. Since seven differences are statistically significant, it seems reasonable to conclude that there is a positive relationship between level of anxiety and frequency of evocation of subjective, personalized, self-centered response tendencies on the Rorschach.

Prediction 3 is also well supported by the data. With but one exception (variability in number of responses) the results are in the predicted direction. Since three of the seven

TABLE 1

COMPARISON OF HIGH AND LOW ANXIOUS SUBJECTS UNDER STANDARD AND "EGO-INVOLVING" CONDITIONS

VARIABLE	HIGH ANXIETY GROUP (N = 20) MEAN OF TRANSFORMED PERCENTAGES	LOW ANXIETY GROUP (N = 20) MEAN OF TRANSFORMED PERCENTAGES	p
No. of card rejections	1.30	2.40	<.50
No. of *W* responses	19.60	27.80	.01
No. of *W* responses	23.28	26.50	.10
No. of *W* and *W* responses	32.24	42.68	.01
No. of *P* responses	34.20	39.70	.02
No. of weighted color responses	14.40	15.10	<.50
No. of irrelevant verbalizations	32.84	31.42	.30
No. of *FM*, *M*, and *m* responses	39.42	31.13	.01
No. of "no form" responses	7.90	2.14	.02
No. of poor form responses	18.70	10.50	.01
No. of form *primary* responses	17.12	22.84	.02
No. of form *secondary* responses	18.68	14.40	.05
No. of sex and anat. responses	15.52	14.20	.30
No. of *FK* and *K* responses	10.32	8.94	.30
No. of perseverative responses	10.52	6.00	.05
No. of *Fc*, *cF*, and *c* responses	10.32	15.94	.02
No. of cards given only one location	32.42	31.48	.40
No. of cards given more than three locations	12.82	8.24	.05
Reaction times *under* 5 seconds	43.68	38.34	.02
Reaction times *over* 15 seconds	17.52	6.48	.01
Total times *under* 25 seconds	32.58	31.16	.30
Total times *over* 75 seconds	12.98	9.62	.10
No. of times moved cards	29.52	28.12	.30

Note.—The respective means and standard deviations were: HA: Mean 25.6 responses, *SD* 11.9 responses; LA: Mean 20.7 responses, *SD* 13.3 responses. Neither the difference between means nor the difference between deviations is significant ($p > .05$).

[5] Some of the intermediary steps between the assumptions and these specific predictions have been omitted. This was because detailed consideration of each prediction would have involved problems beyond the scope of an article of this kind. However, in a later publication, one of the writers (S.B.S.) intends to discuss the intermediary steps.

differences are statistically significant, and one other difference would occur only ten times in a hundred by chance, it can be concluded that there is a positive relationship between level of anxiety and variability of performance on certain Rorschach indices. The evidence is also consistent with the assumption that this variability in performance can be deduced from a postulated positive relationship between level of anxiety and strength of defense against this anxiety when the need for defense is evoked.

Two additional predictions were made:

Prediction 4 (*Hypothesis 3*). From assumptions 1, 2, and 3, it is predicted that, under standard conditions, the differences between HA and LA *Ss* cited in Predictions 1, 2, and 3 will be significant for responses to the first four Rorschach cards, but these differences are expected to be reduced on the last six cards.

This prediction was tested by comparing performances of HA and LA *Ss* on these two groups of cards in the manner described above. The results can be summarized as follows:

a. There were insufficient cases to consider results on four indices: no form intended responses; number of cards given more than three locations; reaction times over 15 seconds; and number of card rejections.

b. If results from all other indices are considered, then 15 of the 19 are in the expected direction.

c. If analysis is confined to those indices that did discriminate significantly between the groups in the validity study, then all ten are in the predicted direction.

d. While eight of the changes were as predicted from a significant to a nonsignificant difference, in no case was this change itself statistically reliable.

Prediction 5 (*Hypothesis 4*). From assumptions 1, 2, and 3, it is predicted that, under "*ego-involving*" conditions, the differences between HA and LA *Ss* cited in Predictions 1, 2, and 3 should be significantly increased.

This prediction was tested by comparing the results of the four subgroups on all indices. Since inspection of these data showed that, in comparison with level of anxiety, the type of instructions had relatively little effect upon performance, it was not necessary to analyze results for all indices. While selection of particular indices is necessarily arbitrary, the most obvious choice is to take two extreme cases: one in which inspection of the data indicates the instructions apparently had little effect, and one in which performance seems to have varied under different conditions of administration. Two other criteria were used in our selection of indices—an insistence on high scores so that the role of chance fluctuations would be minimized, and a demand that the indices discriminate significantly between the two groups under standard conditions; that is, they had to be valid measures.

The two anxiety indicators selected were number of whole and cut-off whole responses, and number of movement (*FM*, *M*, and *m*) responses. Since the distributions of individual percentage scores on these indices were skewed and since very few individual scores reached 50 per cent, it was again necessary to transform these percentages in the manner described above. Analysis of variance of these data revealed that in neither case was there a significant interaction between level of anxiety and type of instructions (with *W* and *Ẅ* responses F was 3.14 with 1 *df*, and with *FM*, *M*, and *m* responses F was 0.29 with 1 *df*). Consequently, it seems reasonable to conclude that the type of instructions did not have a significant effect upon performance.

DISCUSSION

From the viewpoint of the efficacy of these Rorschach anxiety signs, both the known reliability and validity of the questionnaire (8, 10, 11, 12) and the internal consistency of our data afford encouraging evidence of the diagnostic efficiency of certain indices. Of particular interest are two controversial findings: the *negative* relationship between level of anxiety and frequency of *surface* shading responses (*Fc*, *cF*, and *c*), and the form secondary results. The former finding is congruent with our analysis (see footnote 1) and at variance with Eichler's data. While further research is clearly indicated, it is interesting to note that, while 16 of the 20 LA *Ss* obtained higher *Fc* scores than the median percentage, only four of the HA *Ss* did, and a chi-square test (corrected for continuity) yields p values between .01 and .02. Exactly the same number (16 vs. 4) of

Ss scored above and below the median percentage with the poor form index, although here the direction of the difference was reversed. Other measures yielding results which suggest that they may be applicable to *individual* diagnosis were number of whole and cut-off whole responses, number of movement (*FM*, *M*, and *m*) responses, and proportions of reaction times under five seconds; in all of these cases 15 of the HA, and only 5 of the LA *Ss* deviated from the median percentage in the direction predicted for the HA group. In all these instances a chi-square test, corrected for continuity, gave p values between .02 and .05. While the *group* results for certain other indices yielded significant differences (see Table 1), more overlapping in individual scores reduced the clinical significance of these findings.

The other controversial finding was obtained with the form secondary (*K*, *C'F*, *C'*, *cF*, *c*, *C*, and *CF*) index. While the group results with this measure were significant, if Wittenborn's (14) formula is used, precisely the opposite result is obtained ($p = .002$, with a two-tailed distribution). The discrepancy between these results is due to the great disparity in the number of whole responses given by our HA and LA *Ss*. Further work is needed before conclusive evidence can be obtained as to the relative diagnostic efficacy of these two indices.

The following picture of the effects of a specific anxious experience upon functioning seems to emerge: self-oriented or self-centered responses are evoked, awareness of the nature of the objective stimulus is impaired, and effectiveness of responsiveness to the external stimulus is reduced.

Apart from supporting the assumptions, the results have once more demonstrated the precision which can be obtained in personality experimentation if *preselected* groups of *Ss* are used. The present writers are convinced that attempts to vary motivation in randomly chosen *Ss* not only yield less precise data, but also, that such results are extremely difficult to interpret. For instance it is highly improbable that all *Ss* respond in the same way to situational pressures and, more significantly, it is virtually impossible, at least at the present time, to obtain any accurate estimate of the degree of efficacy of such experimentally induced states. If "ego-involved" instructions are given to randomly chosen *Ss*, and if, as is quite likely, both high and low anxious individuals are selected, then their different reactions to such instructions (10) may easily result in a "cancelling out" of a psychologically significant difference. Finally, if induced motivations are used in attempts to test the validity of projective tests (4), it is only too apparent that the conditions created can be so different from those under which clinical tests are normally administered that considerable doubt is cast upon the generality of such experimental findings.

SUMMARY

1. This study was designed to test the validity of certain Rorschach anxiety indices which have been shown (4) to reflect the effects of exposure to a stressful situation, and to examine the generality of certain hypotheses derived from a theoretical framework.

2. Two groups of 20 high and 20 low anxious college students were selected by means of a validated and reliable (8, 10, 11, 12) test anxiety questionnaire. Half of the subjects from each anxiety group were given the Rorschach test under standard conditions, and half were given allegedly "ego-involving" instructions.

3. The results provide encouraging evidence as to the validity of certain anxiety indicators and support the theoretical assumptions from which most of them were derived. The data are also consistent with a hypothesis about the effects of "learning" (adaptation to the test situation) on Rorschach performance, but these results lack statistical reliability. Similarly, although there are some differences in performance under standard and allegedly "ego-involving" conditions, these results are not significant.

4. Finally, it seems reasonable to argue that these data are sufficiently consistent to justify further research on the Rorschach and other projective tests with *preselected* groups of high and low anxious subjects.

REFERENCES

1. BECK, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
2. BENTON, A. L. The experimental validation of the Rorschach test. *Brit. J. med. Psychol.*, 1950, **22**, 45–58.

3. CRONBACH, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.*, 1949, **46**, 393–429.
4. EICHLER, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344–355.
5. FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural and medical research.* (3rd Ed.) New York: Haffner, 1948.
6. HERTZ, MARGUERITE R. *Frequency tables for scoring responses to the Rorschach inkblot test.* (3rd Ed.) Cleveland, Ohio: Western Reserve Univer. Press, 1951.
7. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique.* Yonkers: World Book Co., 1942.
8. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 161–173.
9. SARASON, S. B. The test situation and the problem of prediction. *J. clin. Psychol.*, 1950, **6**, 387–392.
10. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
11. SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.
12. SARASON, S. B., & GORDON, E. M. The test anxiety questionnaire: scoring norms. *J. abnorm. soc. Psychol.*, 1953, **48**, 447–448.
13. SNEDECOR, G. W. *Statistical methods.* (4th Ed.) Ames, Ia.: Iowa State Coll. Press, 1946.
14. WITTENBORN, J. R., & METTLER, F. A. A lack of perceptual control score for the Rorschach test. *J. clin. Psychol.*, 1951, **7**, 331–334.

Received April 2, 1953.

RELIGIOUS BELIEFS OF CATHOLIC COLLEGE STUDENTS AND THEIR ATTITUDES TOWARD MINORITIES

CHARLES T. O'REILLY AND EDWARD J. O'REILLY
University of Notre Dame *DePaul University*

THE relationship between religion and prejudice is not clear. Some investigators, among them Rosenblith (11), Turbeville and Hyde (14), Brown and Lowe (3), and Gough (7), have reported that people who acknowledge church membership tend to be less tolerant than those who do not. Other investigators like Parry (10), Harlan (8), Bettelheim and Janowitz (2), and others have found either slight differences or no differences between the attitudes of members of different religions. Most studies can only relate attitudes to church membership because they only secure information about the subject's religious preference, and by itself this tells us nothing about the subject's "religiosity."

Evans (4) found that religion played an important role in the lives of those who fell into the high or the low quartile of the Levinson and Sanford (9) scale for the measurement of anti-Semitism. His suggested explanation for this was that the low in prejudice adhere to a humanitarian, positive religion, while the high in prejudice follow a narrow religion which tends to foster ethnocentrism. This explanation may do something to clarify the relationship between religion and prejudice and help to explain why people who admit church membership often reveal themselves as prejudiced, thus calling into question the traditional conception of religion as fostering values like brotherhood and tolerance.

The study reported here does not attempt to distinguish between the humanitarian and the narrow types of religion. It is concerned with the attitudes toward Jews and Negroes held by Catholic college students, and the relationship between these attitudes and their attitudes toward certain teachings of the Catholic Church. Important differences were found in the religious beliefs of people who acknowledged membership in the Catholic Church, as well as the fact that these differences in beliefs were related to differences in attitudes toward minorities. It follows, then, that the mere statement of church preference is not enough in order to study the relationship between religion and attitude.

METHOD

As part of an investigation of the attitudes of Catholic students, students in Catholic colleges in three sections of the United States answered a questionnaire which contained, among other things, the 52-item Levinson and Sanford (9) anti-Semitism scale; a 16-item scale to measure attitude toward the Negro; and a 15-item scale to measure the extent to which the subjects (*S*s) agreed with the Catholic Church on certain social, moral, and religious questions. This religion scale was adapted from one developed by Fichter (6). Material obtained from two of the groups studied is reported here.

The M group included 92 white males attending a Catholic, liberal arts college for men in a large southern city. All these *S*s were seniors. The O group included 120 white females attending a Catholic, liberal arts college for women in the same city. They were almost evenly divided into juniors and seniors.

Two men did not answer questions about attendance at Mass and about reception of Holy Communion during the Easter season, which were designed to determine whether the *S*s were practicing Catholics. One woman answered these questions negatively. All *S*s, however, identified themselves as Catholics, and, with the above exceptions, all of them regularly attended church services.

The anti-Semitism scale had a split-half reliability of .92 raised to .96 by the Spearman-Brown formula. The anti-Negro scale had a split-half reliability of .93 raised to .96, and the religion scale a reliability of .90 raised to .95. When tested with another group, the religion scale had a correlation of .60 with the Thurstone and Chave (13) scale for measuring attitude toward the church. The correlation between the anti-Semitism and anti-Negro scale was .69 for the women and .68 for the men; and the correlation between the anti-Semitism and religion scale was .39 for the women and .31 for the men.

RESULTS

Mean scale scores, etc. are shown in Table 1. The *S*s were divided into those whose scores fell into the high and the low halves of the religion scale. Seventy-one per cent of those scoring high on the religion scale also were in the high scoring half of the anti-Negro scale; while 71 per cent of those who scored low on the religion scale also scored low on the anti-Negro scale ($\chi^2 = 39.92$; $p < .01$). Sixty per cent of those scoring high on the religion scale also scored high on the anti-Semitism scale; while 61 per cent of those scoring low on the

TABLE 1

MEAN ANTI-SEMITISM, ANTI-NEGRO, AND RELIGION SCORES, AND *SD*'S, O AND M GROUPS*

GROUP	*N*	ANTI-SEMITISM		ANTI-NEGRO		RELIGION	
		MEAN	*SD*	MEAN	*SD*	MEAN	*SD*
O	120	143.12	44.8	49.3	21.3	41.9	12.9
M	92	161.57	45.4	57.9	20.5	48.5	13.6

* All statements in each scale were negative. Complete agreement with a statement indicated an unfavorable attitude toward a minority or toward the teachings of the church, and received a score of 7.0. Complete disagreement with a statement indicated a favorable attitude and received a score of 1.0. Hence, low scores indicate a generally favorable or unprejudiced attitude toward a minority, or a favorable attitude toward the church, and vice versa. Possible range in the anti-Semitism scale was 52–364. The O range, 56–278; M range, 64–281. The possible anti-Negro range was 16–112. The O range, 16–107; M range, 16–95. The possible religion scale range was 15–105. The O range, 18–78; M range, 15–86.

religion scale also scored low on the anti-Semitism scale ($\chi^2 = 9.99$; $p < .01$).

Nearly all the *S*s said that religion had either a great or a moderate influence upon their upbringing. In an Allport and Kramer (1) study, 28 per cent, and in Rosenblith's (11) study, 34 per cent of the *S*s said that religion had a very marked influence upon their upbringing. The "very marked" category in these two studies corresponds to the "great" category in the present study, which accounted for 74.5 per cent of the *S*s. The higher percentage of Catholics reporting a great or very marked influence of religion upon their upbringing seems to confirm the findings of Ferguson (5), Thurstone and Chave (12), and others who found that Catholics were more influenced by religion than most Protestants and Jews.

When Allport and Kramer combined the A and B categories found in Table 2, they found that a significantly ($p < .01$) larger number of those in the combined group fell into the more prejudiced group. Rosenblith did not find such a relationship in her study, but she did find that the number who reported no religious influence was significantly more in the less prejudiced or lower scoring half in regard to anti-Semitism. Only about 2 per cent of the *S*s in the present study fell into the C and D categories, and no such tendency was observable.

Religion had greatly influenced 74.5 per cent of the *S*s, and 47 per cent of these people fell into the high scoring half of the religion scale, while 59 per cent of those who said that they were moderately influenced by religion fell into the high scoring half of the religion scale ($\chi^2 = 2.28$; $p > .05$). Those who reported that religion had greatly influenced their upbringing had 53 per cent of their number in the more prejudiced half of the anti-Negro scale and 52.5 per cent in the more prejudiced half of the anti-Semitism scale (not significant).

TABLE 2

PERCENTAGES OF SUBJECTS, VARIOUS STUDIES, BY DEGREES OF INFLUENCE OF RELIGION ON UPBRINGING

DEGREE OF INFLUENCE	ALLPORT & KRAMER (1)	ROSENBLITH (11)	O AND M
A. Very marked	28.0	34.0	74.5
B. Moderate	41.0	47.0	23.1
C. Slight	21.0	13.0	1.8
D. None	10.0	5.0	.6

When asked whether Negro Catholics should have their own parishes, 46 per cent of the men and 56 per cent of the women opposed segregation.[1] Those who believed that Negroes should have their own parishes had 76.5 per cent of their number in the high scoring half of the religion scale; while those who were against segregation had 74.5 per cent of their number in the low scoring half of the religion scale ($\chi^2 = 55.09$; $p < .01$). Those favoring segregation had 66.6 per cent of their number in the more prejudiced half of the anti-Negro scale; while 65.4 per cent of those opposing segregation were in the less prejudiced half of the anti-Negro scale ($\chi^2 = 21.82$; $p < .01$).

SUMMARY AND CONCLUSIONS

The anti-Semitism and anti-Negro scores of 92 white, Catholic, male college students and 120 white, Catholic, female college students, who were attending Catholic colleges in a southern city, were related to their scores on a religion scale. The *S*s' estimation of the influence of religion upon their upbringing was not significantly related to scores on the religion scale, or to anti-Semitism and anti-

[1] In two middle western groups, 82 per cent of the men and 69 per cent of the women opposed segregation in parishes, indicating a possible regional difference.

Negro scores. There was a very significant tendency, however, for those scoring high on the religion scale to be less favorable toward Jews and Negroes and to favor the segregation of Negroes in their own parishes, while those who scored low on the religion scale were significantly less prejudiced and were opposed to segregation.

In view of the differences in religious attitudes among these students, all of whom listed themselves as Catholics, and because of the relationship between their religious attitudes and their attitudes toward minorities, it is evident that a mere statement of religious preference or church membership is not an adequate criterion to use when investigating attitudes toward minorities.

REFERENCES

1. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9–39.
2. BETTELHEIM, B., & JANOWITZ, M. *Dynamics of prejudice.* New York: Harper, 1950.
3. BROWN, D. C., & LOWE, W. L. Religious beliefs and personality characteristics of college students. *J. soc. Psychol.*, 1951, **33**, 103–129.
4. EVANS, R. I. Personal values as factors in anti-Semitism. *J. abnorm. soc. Psychol.*, 1952, **47**, 749–756.
5. FERGUSON, L. W. Socio-psychological correlates of the primary attitude scales: I. Religionism. *J. soc. Psychol.*, 1944, **19**, 81–98.
6. FICHTER, J. H. *Southern parish.* Vol. I. *Dynamics of a city parish.* Chicago: Univer. of Chicago Press, 1951.
7. GOUGH, H. G. Studies in social intolerance: IV. Related social attitudes. *J. soc. Psychol.*, 1951, **33**, 263–269.
8. HARLAN, H. H. Some factors affecting attitude toward Jews. *Amer. sociol. Rev.*, 1942, **7**, 816–827.
9. LEVINSON, D. J., & SANFORD, R. N. A scale for the measurement of anti-Semitism. *J. Psychol.*, 1944, **17**, 339–370.
10. PARRY, H. J. Protestants, Catholics and prejudice. *Int. J. Opin. Attit. Res.*, 1949, **3**, 205–213.
11. ROSENBLITH, JUDY F. A replication of "some roots of prejudice." *J. abnorm. soc. Psychol.*, 1949, **44**, 470–489.
12. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude.* Chicago: Univer. of Chicago Press, 1929.
13. THURSTONE, L. L., & CHAVE, E. J. *Instructions for using the attitude scale.* Chicago: Univer. of Chicago Press, 1930.
14. TURBEVILLE, G., & HYDE, R. E. A selected sample of attitudes of Louisiana State University students toward the Negro: A study in public opinion. *Soc. Forces*, 1946, **24**, 447–450.

Received April 11, 1953.

ASSUMED SIMILARITY MEASURES AS PREDICTORS OF TEAM EFFECTIVENESS[1]

FRED E. FIEDLER

University of Illinois

SMALL groups of individuals, working as teams, committees, or crews, conduct the overwhelming proportion of the nation's civic, industrial, and military business. The effectiveness with which these groups function is thus of practical as well as of theoretical concern in our society.

This paper reports a series of related investigations which deal with interpersonal relationships within a team as correlates of the team's total effectiveness.

Certain aspects of interpersonal perception have been investigated in previously published studies on therapeutic relationships (3) and interpersonal relations in a social group (5). These studies present techniques for obtaining the so-called "Assumed Similarity (*AS*) scores," which appear to be correlates of liking and warmth in interpersonal relationships. The present series of investigations relates these interpersonal perception variables to team effectiveness.

A first, frankly exploratory study used 14 high school basketball teams. The most promising measures which emerged from this study were then validated on a second group of 12 high school basketball teams (6). The hypotheses derived from both studies were later tested on teams which differed markedly from the first group, namely, three- and four-man surveying parties (4).

THE FIRST BASKETBALL STUDY

High school basketball squads are composed of from 9 to 18 players. These are chosen by the coach from a larger pool of interested boys competing for places on the first team. This study was conducted in the Midwest, where basketball is of considerable importance to high school players and coaches and where a large number of teams can be found.

The Instrument

The tests used in the basketball study were forced-choice questionnaires. These questionnaires consisted of 100 descriptive statements grouped into 20 blocks of five statements each. An attempt was made to construct the blocks so that, in the main, statements within each block would be equally acceptable to the subjects (*Ss*), but descriptive of different personality dimensions.

One block of statements is given as an example.

	Most	Least
1*a*. I find it easy to understand what others are trying to tell me.....		
b. People think I am a hard worker.	X	
c. I don't mind losing my temper when provoked.....		
d. I like people who don't worry about me.....		X
e. People often look to me for leadership.....		

In a self-description *S* would answer these statements by making an X in the left square opposite the statement which *S* considered *most* characteristic of himself, and an X in the right square opposite the statement he considered to be *least* characteristic of himself.

Test procedure and instructions. So-called interpersonal perception scores were obtained by giving each *S* successively three separate questionnaires containing the same blocks of statements. Players were instructed as follows: (*a*) "describe yourself," (*b*) "predict how the person with whom you can cooperate best will describe himself," and (*c*) "predict how the person with whom you can cooperate least well will describe himself."

In addition to these interpersonal perception tests, players named the three team members—not necessarily the best players—with whom they could cooperate best, and the three with whom they could cooperate least well during games.[2]

Interpersonal Perception Scores

Conventional tests are scored by comparing *S*'s response with the "right" response of a key. Our scores are obtained by comparing two questionnaires of the same person. This comparison yields the so-

[1] The research reported here was carried out under Contract N6ori-07135 between the University of Illinois and the Office of Naval Research. The investigations were conducted by a group in which several members of the project staff played a part during various phases. In particular, the writer wishes to express his appreciation to Dr. L. J. Cronbach, then director of the project, and to Messrs. Walter Hartmann, S. A. Rudin, Mrs. Mary E. Ehart, Miss Vivian C. McCraven, and Dr. W. G. Warrington.

[2] This paper discusses only procedures pertinent to the present hypothesis. A detailed account of the studies is presented in (2, 6, 7, 9).

TABLE 1

CORRELATIONS BETWEEN THE DEC. 31, 1951 CRITERION AND INTERPERSONAL PERCEPTION SCORES IN 14 BASKETBALL TEAMS

ASSUMED SIMILARITY MEASURE	CORRELATION (rho)	
	MEDIAN SCORE IN TEAM AND CRITERION	SCORE OF MOST PREFERRED CO-WORKER AND CRITERION
ASp	−.25	−.63*
ASo	−.03	−.69**

* $p < .05$ according to Olds's tables (11).

** $p < .01$ according to Olds's tables (11).

The large number of exploratory tests run on this first sample does not allow interpretation of significance levels. These are here given only as a point of reference.

called "Assumed Similarity" measures listed immediately below. Tentative interpretations are provided which are based in part on evidence from previous studies.

Assumed Similarity Scores

1. *ASp*—a measure of assumed similarity obtained by comparing (correlating) the *S*'s self-description with his prediction of the self-description of his *positive* choice, the best co-worker. High *ASp* appears to be related to personal liking and perhaps warmth for the chosen person (3, 5).[3]

2. *ASn*—a measure obtained by comparing the *S*'s self-description with his prediction of his *negative* choice. A high *ASn* score may, on the basis of the interpretation above, indicate a feeling of personal closeness and warmth for the negative choice.

3. *ASo*—a measure obtained by comparing *S*'s prediction for his positive choice with his prediction for his negative choice. This measure is interpreted as "set" to differentiate between people. Since *ASn* and *ASo* are highly correlated, we will concern ourselves here only with the scores *ASp* and *ASo*.

The Criterion

Group effectiveness was measured by proportion of games a team had won. The date used here was December 31, 1951,[4] before many changes in team personnel had taken place and by which time each team had played from 8 to 12 games.

In general, small schools are handicapped by having relatively few eligible students. However, teams generally compete with other neighboring schools of comparable size, thus equalizing some of the differences which would favor large schools. The criterion reliability was estimated by comparing the proportion of games a team had won during the first and second halves of the season. At this time it was possible only to estimate reliability for the second sample which was tested with end-of-season criteria. The corrected criterion reliability estimate for these 12 teams was .88. The corresponding reliability for the December 31 date could not be computed because too few games had been played, but it is undoubtedly lower (7).

Relation of Assumed Similarity Measures to Basketball Criteria

Our hypothesis states that interpersonal perception scores are related to the proportion of games a team wins. We tested whether team members of an effective team will, on the average, perceive each other differently from members of ineffective teams. Here we correlated the team median of the *AS* scores with the criterion. As can be seen from the second column in Table 1, correlations between the criterion and median scores are generally near zero.

The group may, however, also express its attitude by the type of person whom most members of a team choose as their best co-worker. In order to get at this attitude we considered only the *AS* scores of those members of the various teams which had received the greatest number of "best co-worker" votes. As can be seen in Table 1, this procedure suggests that the *AS* scores of the team's "most preferred co-worker" correlate with the criterion in the negative direction.

In previous studies, high *ASp* seemed to be related to warm, empathic interpersonal relationships. We expected to find these relationships—hence also high *AS*—to be prevalent in effective teams. The present findings are thus in the direction opposite to that which was anticipated. They suggest that the most preferred cooperators in effec-

[3] *AS* scores can be computed as correlations or as difference scores. To avoid possible confusion we will in this paper speak of high *AS* as meaning high assumed similarity, i.e., a high correlation between either the self and the predicted self of another person, or between the predicted self-descriptions of two other persons. In terms of *D* scores, this would imply small differences between the two descriptions on which *AS* is based.

[4] The original report (5) deals with December 15 as the criterion date. Since some teams had at that time played fewer than 4 games, the December 31 date appears to be a better criterion estimate.

TABLE 2

SECOND STUDY: POINT BISERIAL CORRELATIONS BETWEEN THE CRITERION AND ASSUMED SIMILARITY SCORES OF MOST PREFERRED CO-WORKER

ASSUMED SIMILARITY MEASURE	$r_{pt.\ bis.}$	t	p
ASp	−.20	.53	
ASo	−.58	2.20	.03

tive teams tend to be somewhat less warm and emotionally less involved with persons whom they choose as work companions than is the case of keymen in less effective teams.

The correlations in Table 1 are, of course, based on a small sample of teams, and on only moderately reliable scores, .62 and .61 for *ASp* and *ASo* respectively (9). In addition, they are the survivors of a considerable number of exploratory measures. A validation attempt, therefore, became essential.

VALIDATION STUDY ON BASKETBALL TEAMS

The second sample of teams was studied solely for the purpose of testing relationships which were significant at the .05 level in the first study, i.e., on measures *ASp* and *ASo* of the most preferred co-worker.

The only major modifications were in the method of choosing teams and testing significance. We selected, toward the end of the season, 9 teams which had had a predominantly winning season and 9 teams which had had a predominantly losing season, and requested their cooperation.[5] These came from the upper and lower third of a roster of over 50 teams. We tested 7 "good" teams and 5 "poor" teams which agreed to cooperate. Since the teams were dichotomously selected, point-biserial correlations were here used to estimate the degree of the relationship. The significance of the difference between the scores of "good" and "poor" teams was tested by the usual t test; inspection of the data indicated that the conditions for applying a t test were not violated. The small samples and the not very high reliability of the scores suggest caution in interpreting these data.

As can be seen from Table 2, the point biserial correlation between the criterion and

[5] The writer gratefully acknowledges the invaluable assistance received from Clyde Knapp and Harry A. Combes of the University of Illinois.

ASp			*ASo*		
Q CORRELATIONS	GOOD TEAMS	POOR TEAMS	Q CORRELATIONS	GOOD TEAMS	POOR TEAMS
.55			.55		
.50			.50		
.45	x	xx	.45		x
.40	x	x	.40		
.35			.35		x
.30	x	x	.30		
.25			.25	x	xx
.20	x		.20	x	x
.15	x	x	.15	x	
.10			.10	x	
.05			.05	x	
.00	x		.00		
−.05			−.05		
−.10			−.10	x	

FIG. 1. *ASo* AND *ASp* OF MOST PREFERRED CO-WORKER PLOTTED AGAINST THE CRITERION IN THE SECOND SAMPLE

ASo of the team's most preferred co-worker is −.58. We attempted to validate two measures. Only one of these (*ASo*) reached significance. We are therefore not justified to consider the relation of *ASo* to team effectiveness in basketball as established. *ASp* of the teams' most preferred co-workers did not correlate significantly with the criterion even though the correlation is in the anticipated direction. We have plotted the measures *ASp* and *ASo* of the most preferred team members from good and poor teams. (See Fig. 1.)

In addition, we also computed validities for the end of the season when all league games had been played. These validities are generally lower than those of December 31. It is clear from these data that *ASp* in contrast to *ASo* is not a promising predictor of team effectiveness (7).

STUDY ON SURVEYOR TEAMS[6]

The Hypothesis

The basketball team studies led to one major hypothesis: *Members of effective teams will prefer co-workers who assume relatively little similarity between the persons whom they choose and those whom they reject as their own co-workers.*

[6] The writer is indebted to Prof. M. O. Schmidt, Civil Engineering Department, University of Illinois, whose interest and cooperation made this phase of the study possible.

Since we interpret high *AS* to be indicative of warmth toward, and acceptance of, others we also explored whether team effectiveness and congeniality are negatively related.

The following study was designed to test this hypothesis, and to obtain evidence regarding the additional question.

Selection of Groups

Student surveying parties work in teams of three to four men. The *S*s were 71 civil engineering students taking a required course in surveying. The course consists of two parts. The first part is taught on the university campus on a full-day basis, lasting three weeks; the second part covers five weeks. This is offered at a university-operated surveying camp in northern Minnesota where students concentrate on field problems in relatively difficult terrain. The camp is almost completely isolated and self-contained. Students as well as faculty members eat, sleep, and work there, and students are under practically continuous supervision of their instructors.

Organization of the course at camp. While at camp, the students were divided into six sections, one instructor remaining in charge of each of the sections throughout the camp period. Each section consisted of three or four parties, and each of the parties consisted of three or four men. A total of 22 surveying parties was formed.

Differences between basketball teams and surveying parties. Obviously basketball teams differ in many respects from surveying parties. The differences which we considered among the most important are the following:

a. Basketball teams require physical coordination, relatively little verbal interaction. Surveying is primarily an intellectual task requiring frequent verbal communication.

b. While basketball squads consist of 9–18 members, the surveying parties in our study were no larger than four men.

c. Basketball teams work under considerable time pressure. Speed in surveying is only of secondary importance.

d. Members of basketball teams are highly identified with the team, and personally involved with their team's success. This identification and involvement is almost completely absent in surveying teams. The students were graded individually, and no benefits were derived from being in a "good" surveying party. This is shown by the fact that none of the students were interested in their instructor's opinion of their *team*.

The Instrument

As in the basketball studies, *S*s responded to three identical questionnaires, predicting a preferred, and predicting a rejected co-worker. Unlike the forced-choice questionnaires used in the basketball study, the surveyor tests consisted of 60 statements, each of which was to be marked on a seven-point scale ranging from *definitely true* to *definitely untrue*. The statements were pretested on a 180-item questionnaire.[7] Statements were selected on the basis of an item analysis to obtain items with large variances on self-descriptions. Statements such as "I am very discriminating in my choice of friends," "I am not likely to admit defeat," or "when a person is a failure it is his own fault," were used. The instrument presents a considerable improvement over the tests used in the basketball studies. The reliabilities for *ASp* and *ASo* are .83 and .93 respectively (2). In addition, the tests require less time for administration.

By comparing the two tests by means of the statistic *D* (1), it is possible to obtain a score indicating how similarly any two of the questionnaires have been marked.

Test procedure and instructions. The instructions and administration of questionnaires followed those of the basketball studies, with a few, relatively minor, exceptions. The *S*s could predict their preferred and least preferred persons from among those with whom they had previously worked. These did not have to be persons at the camp. The *S*s again completed sociometrics regarding the three persons within their section (10–15 *S*s) whom they personally liked most and liked least. They similarly named three *S*s whom they preferred most, and the three whom they preferred least as co-workers.

The Criteria of Effectiveness

Instructors were asked to rank all teams in their section in terms of the following:

a. Accuracy with which surveying jobs were done by various parties.

b. Speed with which the jobs were done.

c. Congeniality of the teams in terms of lack of conflict and smooth-running field operations.

In addition, students in all sections were asked to "rank all parties in the section from best to poorest." This constitutes our *students' ratings* criterion.[8]

Accuracy is the main criterion in surveying. It was, therefore, the only criterion on which we attempted to validate the hypothesis derived from the basketball study. The one-tailed test of significance was applied, therefore, to the accuracy criterion only. Tests relating to other criteria were exploratory.

[7] We are indebted to Col. R. W. Faubion, Commanding Officer, Det. #3, Human Resources Research Center, Chanute Air Force Base, for permission to pretest this instrument.

[8] Students' ratings of Section V could not be used. The *S*s in that section had been in more than one surveying party, and a number of students rated teams other than the main teams rated by the instructor of that section.

Each instructor could rank only the three or four surveying parties in his own section. Ranking of parties from different sections was standardized to permit comparison of all parties. *AS* scores for all 22 teams could then be correlated with the various criteria.

The fact that instructors' frames of reference differ decreases to some extent the criterion reliability. This would tend to obscure any relationships present, and it would increase the probability of accepting the null hypothesis when a true difference exists. Table 3 presents the intercorrelations of the four criteria used in this study for three- and four-man parties.

TABLE 3

INTERCORRELATIONS (r) OF CRITERIA FOR THREE- AND FOUR-MAN SURVEYOR TEAMS*

	ACCURACY	SPEED	CONGENIALITY RATINGS
Speed	.79		
Congeniality ratings	.15	.52	
Student ratings	−.34	.15	.39

* (See footnote 8.) Based on N's of 22, except correlations with the student rating criterion in which N's = 18.

TABLE 4

COMPARISON OF *AS* OF PREFERRED CO-WORKERS IN TEAMS RATED HIGHEST AND LOWEST IN ACCURACY (N = 12)

INTERPERSONAL PERCEPTION SCORES	MEAN OF HIGHEST TEAMS*	MEAN OF POOREST TEAMS	t	p	r (N=22)
ASp	12.96	12.24	.56		
ASo	20.61	15.32	3.30	.025	−.51

* In terms of D's. A high score indicates *low* assumed similarity.

Relation of Assumed Similarity Measures to Criteria in Surveying

The accuracy criterion. Our major hypothesis states that the assumed similarity of most preferred co-workers in good teams will be relatively low. The preferred co-workers in relatively poor teams will have high *ASo* scores.

Our population of teams consists of 22 surveying parties, divided into six different sections. This division presents difficulties in statistical treatment of the data since no criterion was available for comparing teams from different sections. We have here tested the hypothesis by two methods.

a. We compare the best and the poorest teams within each of the six sections. We can then ask whether the *AS* scores of the preferred co-worker in the best team from each section are lower than the *AS* scores of the preferred co-worker in the poorest team in each section. Since the two teams for each section are evaluated by the same instructor, the matched t test can here be used. This does, however, reduce to 12 the number of teams (cases) used in the analysis. (See Table 4.) As can be seen in Table 4, *ASo* differences are significant. *ASp* shows only a negligible difference.

b. A somewhat more satisfactory indication of the *degree* of relation comes from a second analysis. Criterion ratings were converted to z scores, and then correlated with the most preferred co-workers' *AS* scores, both *ASp* and *ASo*.

Table 4 also presents the r's between the primary criterion, accuracy, and these two *AS* scores. As can be seen, the hypothesized relationship has been found between the criterion and *ASo* of the most preferred co-worker. Hence, persons chosen as most preferred co-workers in effective (i.e., accurate) teams perceive a greater difference between those whom they prefer and those whom they reject as co-workers than keymen in less accurate teams. The findings thus are consistent with the hypothesis induced from the results which were obtained in the study on basketball teams.

The Secondary Criteria

No significant relation was found between *AS* scores and secondary criteria. The relation between *AS* scores on students' ratings is, however, in the opposite direction of those found for accuracy and other criteria based on instructors' ratings. In other words, the students tended to rate those teams as better in which the preferred co-worker assumed relatively high similarity to his negative choice. This appears to lend some support to the interpretation that preference for a person with high *AS* is related positively to congeniality within the team. (See Table 3.)

The relation of congeniality and effectiveness was explored by one further step. An Intrateam Preference Index (*IPI*), defined as an index of congeniality, was devised and correlated with each of our four criterion ratings.

This measure of congeniality is based on the following considerations:

a. Each person had rated the other 10–15 members of his section in terms of how well he liked them as co-workers.

b. A subject worked in a three- or four-man team. He could choose his preferred co-workers within his own three- or four-man team, or he could prefer others in his section who were not in the team.

c. We assume that a team whose members choose one another is more congenial than one whose members reject one another or choose outsiders.

The measure is computed by the formula:

$$\text{Intrateam Preference Index } (IPI) = \frac{\left(\begin{matrix}\text{choices}\\ \text{within}\end{matrix} + \begin{matrix}\text{rejections}\\ \text{without}\end{matrix}\right) - \left(\begin{matrix}\text{choices}\\ \text{without}\end{matrix} + \begin{matrix}\text{rejections}\\ \text{within}\end{matrix}\right)}{nk - n}$$

n = number of men within the team
k = number of choices made by each individual

When this intrateam preference index (*IPI*) was correlated with our four criterion ratings, the following relationships appeared: The correlation between accuracy and the *IPI* was negative (−.23), while other criterion ratings correlated positively with the *IPI*. The highest correlation was found between students' ratings and the *IPI* (.37). When we compare the most preferred co-worker's *ASo* scores of teams considered best and those considered poorest by students, we find higher *ASo* for teams rated high by students and lower *ASo* scores (greater perceived difference between most and least preferred work companions) for teams which students rated as poor. This relationship is not statistically significant ($t = 2.24$; $t = 2.57$ is required for significance of .05). The direction would tend to indicate, however, along with our other data, that effectiveness and congeniality may be inversely related in informal teams. This finding is in accord with the findings in the preceding basketball study. Moreover, it is supported by the results in Halpin's recent study of air crew leaders (8) and a paper by Schachter, Ellertson, and McBride on experimentally assembled student groups (10). The results are not in agreement with Van Zelst's study on construction workers (11). All in all, these data suggest that further study is needed to determine whether or not effectiveness and congeniality are inversely related.

Discussion

Two interpersonal perception scores on assumed similarity were correlated with the criterion in each study. One relation reaches the prescribed significance level every time. The studies thus support the hypothesis that the interpersonal perception variable *ASo* plays a part in group effectiveness.

We feel that the present findings serve primarily to emphasize that research on interpersonal perception in task groups is a fruitful area for continued efforts. This discussion will, therefore, be largely concerned with the implications of these findings for further research.

Let us first examine the measure which yielded significant results.

ASo. This score was obtained by comparing *S*'s prediction of his positive choice for work companion with his prediction of his negative choice. It is thus the similarity which *S* assumes to exist between the person with whom he says he can, and the one with whom he says he cannot cooperate. According to our data, the most preferred co-worker in effective teams tends to perceive these two persons as relatively dissimilar. On the other hand, the most preferred co-worker in ineffective teams tends to perceive these persons as relatively similar. Low *ASo* (i.e., large perceived difference between most and least preferred work companions) may, therefore, reflect an evaluative, critical attitude toward others, as contrasted with warm, empathic interpersonal relations. (Further research is needed to clarify the meaning of this measure.) We have found that *AS* tends to correlate positively with reputed therapeutic competence (3); in subsequent studies it was shown that *S*s assume significantly more similarity to a person who is liked than to someone who is relatively disliked (5, 6). Since *ASo* and *ASn* (*AS* to the least preferred) are highly correlated, we believe that the person with low *ASo* is relatively

unaccepting, and perhaps rejecting, to the person who is not a good work companion.

The most preferred co-worker. While *ASo* in key persons appears to measure relevant factors in team effectiveness, it also points to a phenomenon which may be of more general theoretical importance. Only the scores of the most preferred co-workers correlate with the criterion. When we correlated the team's median *ASo* with the criterion, no significant relationships were found. At present we are inclined to take these results as an indication that members of effective teams use a basis different from that of members of ineffective teams for choosing and rejecting others as cooperators. This interpretation is supported by the positive correlations between *ASo* of the most and of the second most preferred co-workers in basketball teams (.63 and .27 for the first and second samples, respectively). In light of our current interpretation of *ASo*, this would mean that members of effective teams prefer highly task-oriented persons as co-workers. Members of relatively ineffective teams list as their most preferred co-workers the more accepting, relationship-oriented team members. *ASo* in the most preferred team worker is thus possibly an indication of the entire team's attitude toward the task, e.g., an index of the team's morale. Whether or not these relations hold in radically different teams, e.g., formally structured groups, permanent crews, etc., remains to be established.

SUMMARY AND CONCLUSIONS

The present investigations test the hypothesis that group effectiveness is related to the interpersonal perceptions which members of the group have toward one another.

Interpersonal perceptions were measured by correlating identical questionnaires which subjects were instructed to fill out (*a*) describing themselves, (*b*) predicting the responses of their preferred co-worker, and (*c*) predicting the responses of their rejected co-worker.

The first studies used 14 high school basketball teams, tested at the beginning of the season. A second sample of 7 "good" and 5 "poor" teams was collected toward the end of the season for the purpose of verifying relations identified in the first study. A third sample consisted of 22 surveying teams.

The criterion of basketball team effectiveness was the proportion of games the teams had won (at midseason in the first sample, two weeks before the end of the season in the second sample). The criterion in the surveyor study was the instructor rating on accuracy. There was no correlation between the criterion and the median of any assumed similarity score within a team. The assumed similarity score, *ASo*, of the team's most preferred work companion was negatively correlated with the criterion in all three samples. The finding supports the hypothesis that *ASo* of the most preferred co-worker in surveying, and possibly also in basketball, is related to team effectiveness.

The interpersonal perception scores of the chosen person are believed to reflect his outlook on other persons and on the task. Low *ASo* is thought to reflect lack of emotional involvement with teammates and task-oriented attitudes. The group which chooses a differentiating person as preferred co-worker is thus likely to be more concerned with effective task performance, and correspondingly more successful. Some evidence suggests that the more effective surveying teams tend to be less congenial than relatively ineffective teams.

As in previous studies, we found that *S*s assumed greater similarity between themselves and their positive, than between themselves and their negative choices.

REFERENCES

1. CRONBACH, L. J., & GLESER, GOLDINE C. *Similarity between persons and related problems of profile analysis.* Urbana, Ill.: Bureau of Research and Service, Univer. of Illinois, 1952. (Mimeo.) (Tech. Rep. No. 2, Contract N6ori-07135.)
2. CRONBACH, L. J., HARTMANN, W., & EHART, MARY E. *Investigation of the character and properties of assumed similarity measures.* Urbana, Ill.: Bureau of Research and Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 7, Contract N6ori-07135.)
3. FIEDLER, F. E. A method of objective quantification of certain counter-transference attitudes. *J. clin. Psychol.*, 1951, **7**, 101–107.
4. FIEDLER, F. E. *Assumed similarity measures as predictors of team effectiveness in surveying.* Urbana, Ill.: Bureau of Research and Service Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 6, Contract N6ori-07135.)
5. FIEDLER, F. E., BLAISDELL, F. J., & WARRINGTON, W. G. Unconscious attitudes as correlates of sociometric choice in a social group. *J. abnorm. soc. Psychol.*, 1952, **4**, 790–796.

6. FIEDLER, F. E., HARTMANN, W., & RUDIN, S. A. *The relationship of interpersonal perception to effectiveness in basketball teams.* Urbana, Ill.: Bureau of Research and Service, Univer. of Illinois, 1952. (Mimeo.) (Suppl., Tech. Rep. No. 3, Contract N6ori-07135.)
7. FIEDLER, F. E., HARTMANN, W., & RUDIN, S. A. *Correction and extension of the relationship of interpersonal perception to effectiveness in basketball teams.* Urbana, Ill.: Bureau of Research and Service, Univer. of Illinois, 1953. (Mimeo.) (Tech. Rep. No. 5, Contract N6ori-07135.)
8. HALPIN, A. W. The relation between the crew's perception of the leadership behavior of airplane commanders and superiors' ratings of their combat performance. Paper read at Amer. Psychol. Ass., Washington, D. C., 1952.
9. RUDIN, S. A., LAZAR, I., EHART, MARY E., & CRONBACH, L. J. *Some empirical studies of the reliability of social perception scores.* Urbana, Ill.: Bureau of Research and Service, Univer. of Illinois, 1952. (Mimeo.) (Tech. Rep. No. 4, Contract N6ori-07135.)
10. SCHACHTER, S., ELLERTSON, N., MCBRIDE, DOROTHY, & GREGORY, DORIS. An experimental study of cohesiveness and productivity. *Hum. Relat.*, 1951, **4**, 229–238.
11. VAN ZELST, R. H. Sociometrically selected work teams increase production. *Personnel Psychol.*, 1952, **3**, 175–185.

Received April 13, 1953.

THE EFFECT OF REWARD ON ADULT IMITATIVE BEHAVIOR[1]

EDGAR H. SCHEIN[2]

Army Medical Service Graduate School

IN THEIR book *Social Learning and Imitation* Miller and Dollard (6) assert that imitative behavior is learned by the same principles as other kinds of behavior, the principles of instrumental conditioning. This assertion is supported by the findings of a series of experiments carried out with animals and children. The study to be reported below attempted to test the Miller and Dollard theory in a somewhat wider context with adult human subjects. Specifically, the purpose of the experiment was (*a*) to determine whether rewarding imitative responses would make such responses more probable, and (*b*) to determine whether such responses, once "learned," would generalize to a new situation.

METHOD

The subjects (*S*s) were 300 Army inductees at a reception center.[3] They participated in the experiment in groups of five shortly after they had completed their preliminary processing. When they arrived at the experimental room each of these groups was told that they were about to take two tests which were somewhat different from the ones they had just completed as part of their normal processing, but which were, nonetheless, part of the regular testing procedure and just as important.

Altogether, 60 groups (300 *S*s) were run according to the experimental design shown in Table 1.

Each row in the table represents the treatment given to a different set of 10 groups. For the experimental groups Test 1 was a learning situation in which imitation was systematically rewarded. In the control groups all conditions were identical except that imitation was *not* systematically rewarded. Test 2 was a test of generalization, administered in an identical way to experimental and control groups. There were three different types of generalization tests, varying in presumed similarity to Test 1, but each set of experimental or control *S*s received only one type.

Throughout the experiment *S*s sat next to each other facing *E*. Each *S* was given one copy of a card which had five groups of four letters and numbers on it.[4] The *S*s were told that the order and magnitude of the letters and numbers in *each* of the groups was determined according to some organizing principle, and that it was their job to discover what these organizing principles were, by learning to match other sets of letters and numbers to one of the sets on their card. The *E* held up 25 cards, one at a time, each with a group of three letters and numbers on it.[5] Each of these groups was ostensibly constructed by the same principle as one of the five groups on *S*s' cards. If the principle used in making the match was correct, *S* would match correctly; if the principle used was wrong, he would make the wrong match and have to try a different principle.

The *S*s gave their answers out loud by calling out the number of the group on their card which they chose as the correct match for the one held up by *E*. They always gave their answers in order, ostensibly to avoid confusing *E*, who was recording them. After all five men had given their answer for any particular problem, *E* would announce the number of the group which was the correct match, and then go on to the next problem.

Actually there were no principles underlying the construction of any of the groups of letters and numbers. The groups were arbitrarily constructed by *E* to allow for any number of principles to be applicable. Any matching of groups could be "correct" or "incorrect," since any of a large number of principles "might" be applicable.

This situation made it possible for *E* to choose arbitrarily *any* answer as the correct one without arousing *S*s' suspicion in any way. Imitation was then "rewarded" by the following procedure: On items 6 to 25 of the first test, *E* chose as correct the answer that the second man in the order of responding gave, no matter what it was. Therefore, the only way a man answering after the No. 2 man could get the correct answer, was by copying the answer of the No. 2 man. If he did this, he would be rewarded by the acknowledgement on the part of *E* that he had obtained the correct answer.

---

[1] The present study was facilitated by the Laboratory of Social Relations, Harvard University. The author wishes to express his appreciation to Professors Gordon W. Allport, Richard L. Solomon, and Frederick Mosteller for their helpful advice in carrying out the study.

[2] First Lieutenant, MSC, USA.

[3] The *S*s were obtained at the Reception Center at Fort Meade, Maryland. I wish to thank the staff of the Center for their cooperation in making it possible to use inductees for the experiment.

[4] The five groups were as follows:

1. A 6 C 8		3. 2 F 9 A		5. 7 H 3 J
2. K A C 9		4. B 1 H 8		

[5] These cards read:

1. B D 7	10. B 3 A	18. F 5 K
2. 4 G 8	11. B K 2	19. 7 E 2
3. 1 F C	12. F D C	20. 5 A 3
4. C 1 E	13. G E 3	21. 4 B A
5. C 5 7	14. K E G	22. 2 5 B
6. 4 3 6	15. 1 5 7	23. 2 8 7
7. D 2 3	16. K 8 2	24. A 7 D
8. 3 4 C	17. A B E	25. 6 1 K
9. 7 E D		

TABLE 1

EXPERIMENTAL DESIGN

GROUPS	ANSWERS ANNOUNCED	ANSWERS NOT ANNOUNCED
Experimental	Test 1 No. 2 rewarded	Test 2a
	Test 1 No. 2 rewarded	Test 2b
	Test 1 No. 2 rewarded	Test 2c
Control	Test 1 random reward	Test 2a
	Test 1 random reward	Test 2b
	Test 1 random reward	Test 2c

On items 1 to 5, *E* chose as correct an answer which no one in the group had given if this was possible. In the control groups the entire procedure was identical except that the correct answers to the problems were assigned at random by *E*.

Each experimental or control group was given the second test immediately following the first. The test material assumed to be most similar to Test 1 consisted of 25 multiple-choice perceptual problems presented on slides. The problems consisted of series completions, mirror images, and matrix completions.[6] The first slide, showing the problem, was flashed for 10 seconds; then a second slide was flashed for 10 seconds with five plausible looking alternatives on it, none of them the correct answer. The test assumed to be intermediate in similarity to Test 1 consisted of 25 dot numerosity judgments. Each slide had from 100 to 300 dots on it and was flashed for five seconds.[7] The *S*s were told that the problem was one of estimating large numbers of objects such as airplanes when these can only be seen for short periods of time. The *S*s simply reported how many dots they thought were on each slide. The test material assumed to be least similar to Test 1 consisted of 25 multiple-choice problems, on each of which the "best" photographs from the standpoint of technical proficiency had to be selected from a group of five in 25 seconds.

Each of these tests again called for verbal responses, and *S*s gave their answers out loud and in the same order as they had on the first test. However, on this second group of tests no announcement was made as to which answers were correct on any of the problems, thus removing any "reward" for imitation. The experimental and control groups were treated in exactly the same way. At the end of the two tests each group of *S*s was given a short questionnaire in an attempt to survey some of their attitudes toward the experimental procedure and toward each other.

[6] Items were selected from Cattell's Culture Free Intelligence Test with permission of the Institute for Personality and Ability Testing, 1608 Coronado Drive, Champaign, Illinois.

[7] The actual numbers of dots on the 25 slides presented were: 160, 160, 150, 140, 100, 300, 240, 220, 200, 200, 180, 170, 160, 160, 150, 140, 120, 100, 300, 240, 220, 200, 200, 180, 170.

This procedure was designed to test the following hypotheses: (*a*) that the No. 2 man in the experimental groups would be imitated more than the No. 2 man in the control groups on Test 1, and that the amount of imitation would rise through the items of the test (with increasing number of rewards); (*b*) that the No. 2 man in the experimental groups would be imitated more than the No. 2 man in the control groups on Test 2, and that the difference between the groups would decrease as the similarity between Test 1 and Test 2 decreased.

Advantages of the experimental situation. The use of Army inductees in an Army setting had several methodological advantages: (*a*) It was possible to reduce uncontrolled prestige factors to a considerable extent. In most cases *S*s did not know each other personally, or at most had an acquaintance of one or two days' standing. Furthermore, since all the inductees were dressed in fatigues, overt status differences were completely eliminated. All formal communication between *S*s which might have served as a basis for prestige was prohibited during the course of the experiment. (*b*) The 300 *S*s constituted a representative cross section of the population of young men with respect to intelligence, education, personality, and social-cultural background. (*c*) It was possible to present the experiment as a routine test situation. The behavior that was called out was therefore untainted by motives such as "pleasing the experimenter," "not trying hard because it was only research," "making the research come out a certain way," or "guarding against that guy (the experimenter) finding out anything about me."

Analysis of the data. An imitative response was defined as any agreement between a man's answer and the answer of some man who had responded before him on that item. Since the No. 1 man in order of responding could not imitate anyone, and the No. 2 man was chosen as the model, most of the results concern the imitative behavior of the No. 3, No. 4, and No. 5 men in each group. Since no systematic difference in the amount of imitation that these different positions showed was observable, we averaged across the positions in all our analyses.

If the different alternatives for each problem were equally probable, imitation could be measured against a chance level in both experimental and control groups. The *S*s did not in fact choose all of the answers equally often, and we therefore used comparisons between experimental and control groups only.

The analysis of dot numerosity judgments was carried out in terms of the absolute distance between the estimates of the different positions without regard to the true numbers of dots on the slides. The actual numbers affected these differences, but the effect was approximately the same in the experimental and control groups, thus making comparisons possible.

RESULTS

Test 1. Figure 1 shows the mean amount of imitation of the No. 2 man by the No. 3, 4, and 5 men (averaged together) over successive groups of five items on Test 1. Each curve is based on 30 groups.

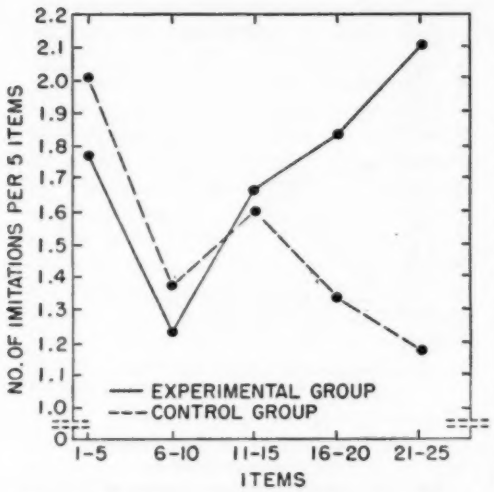

FIG. 1. AMOUNT OF IMITATION THROUGH THE ITEMS OF TEST 1

Each point on a curve shows the amount that the No. 3, 4, and 5 men imitated the No. 2 man on that group of five items.

TABLE 2

FREQUENCY DISTRIBUTIONS OF IMITATION ON ITEMS 16 TO 25 OF TEST 1

NUMBER OF IMITATIVE RESPONSES*	NUMBER OF EXPERIMENTAL SUBJECTS (N = 90)	NUMBER OF CONTROL SUBJECTS (N = 90)
0	2	11
1	7	13
2	17	27
3	10	14
4	21	13
5	14	9
6	10	2
7	6	1
8	1	0
9	0	0
10	2	0

* Mean number of imitative responses for the experimental *Ss* was 3.94 (SD = 2.02); for the control *Ss* the mean was 2.50 (SD = 1.63).

The initial high level of imitation in both experimental and control groups is probably due to the fact that all *Ss* tended to use the same simple hypotheses for matching at the beginning of the test. Since none of the answers in the experimental groups and very few of the answers in the control groups were acknowledged by *E* as correct, *Ss* probably began to use more varied and complex hypotheses, thus lowering the level of imitation on items 6 to 10. By the time items 11 to 15 were reached, *Ss* in the experimental groups probably noticed that the No. 2 man was consistently correct, that he had "caught on" to the principles. The results from that point on show a consistent rise in amount of imitation in the experimental groups, and a consistent decrease in amount of imitation in the control groups. The difference between the experimental and control means for items 16 to 25 is significant at the .001 level (one-tailed t test).

The first hypothesis is thus strongly supported, but several additional findings should be taken into account in making an unambiguous interpretation of the data. (*a*) Individual differences were large (see Table 2). (*b*) The *Ss* were highly inconsistent in their behavior from one item to the next. Very few men settled down to a pattern of either imitating or not imitating. (*c*) The level of imitation reached in the experimental groups was not very high in an absolute sense. If *Ss* had "learned" to imitate, one might expect an average level somewhat higher than 2.1 imitative responses on the last five trials.[8]

Test 2a. Figure 2 shows the results for the generalization test most similar to Test 1. Through all the 25 items the No. 3, 4, and 5 men (averaged together) in the experimental groups imitated the No. 2 man more than the comparable men in the control groups. The difference between the means for all 25 items taken together is significant at the .01 level (one-tailed t test), thus showing clear-cut positive generalization of the imitative response.

Individual differences in responding continued to be large, and *Ss* remained inconsistent in their behavior. The rank-order correlation between the amount of imitation on items 16 to 25 of Test 1 and the amount of imitation on Test 2a is only +.21, indicating that group results cannot be used to infer *Ss*' individual reactions. Those *Ss* who imitated on Test 1 did not always imitate on Test 2a and vice versa.

Test 2b. Figure 3 shows the results for the generalization test of intermediate similarity. The large fluctuations in the curves reflect the

[8] It was not possible from our results to determine what might have happened to the curve of imitation if the test had contained more items.

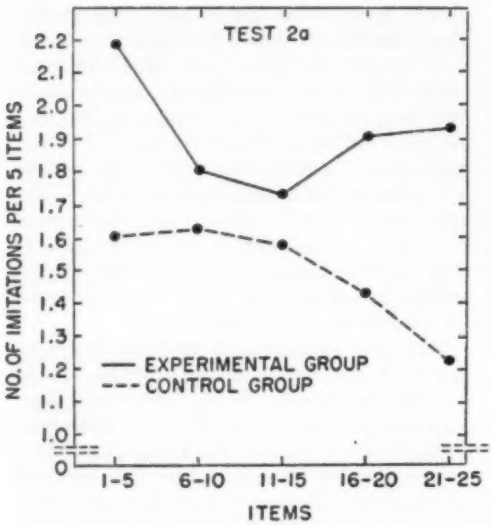

FIG. 2. AMOUNT OF IMITATION THROUGH THE ITEMS OF TEST 2a

Each point on a curve shows the amount that the No. 3, 4, and 5 men imitated the No. 2 man on that group of five items.

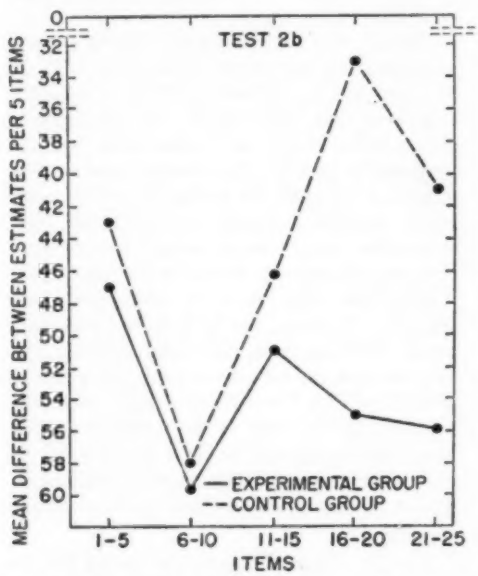

FIG. 3. DEGREE OF AGREEMENT BETWEEN ESTIMATES OF NUMEROSITY THROUGH THE ITEMS OF TEST 2b

Each point on a curve shows the degree to which the No. 3, 4, and 5 men agreed with the No. 2 man on that group of five items.

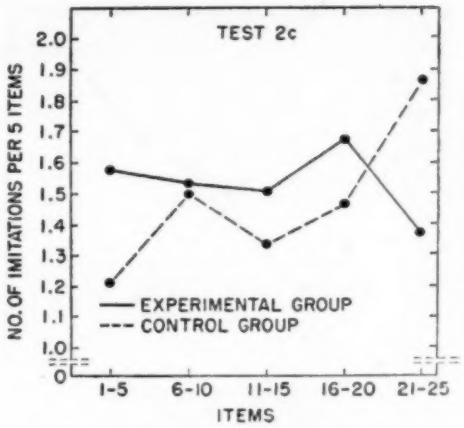

FIG. 4. AMOUNT OF IMITATION THROUGH THE ITEMS OF TEST 2c

Each point on a curve shows the amount that the No. 3, 4, and 5 men imitated the No. 2 man on that group of five items.

effects of the actual numbers of dots on the slides. Slides with large numbers of dots tended to produce uniformly larger differences between the estimates of positions than slides with small numbers of dots on them. On this test a rather surprising reversal occurred. The agreement between the estimates of the No. 2 man and the average of the No. 3, 4, and 5 men in the experimental groups was *lower* than the agreement between these positions in the control groups. There was, in fact, a tendency for *all* positions to agree *less* with each other in the experimental than in the control groups. The difference between the means for all 25 items together is significant at the .05 level (two-tailed t test, since the results are the opposite of what was predicted).

Individual differences and inconsistencies were again large. The rank-order correlation between imitation on items 16 to 25 of Test 1 and agreement on Test 2b was +.14.

Test 2c. Figure 4 shows the results for the generalization test of least similarity. The mean imitation in the experimental groups was slightly higher than the mean in the control groups, but the difference is not significant and the curves cross. The rank-order correlation between items 16 to 25 of Tests 1 and 2c was +.18.

Further findings. Data obtained from the questionnaires indicated that *E*'s judgment of

the relative similarity of the generalization tests to the first test did not agree with *Ss*' perceptions. We asked *Ss* to underline which of the following alternatives they considered to be "most important" in getting a high score on each of the tests: (*a*) ability to concentrate, (*b*) ability to think quickly, (*c*) mostly luck, (*d*) intelligence, or (*e*) previous experience with tests (or practice in judging dots or photographs in the case of Tests 2b and 2c).

Table 3 shows the percentage of *Ss* choosing the different categories on all the tests. It can be seen that Test 2b was thought to be primarily a test of practice in judging, and less a test of intellectual ability than Tests 1, 2a, or 2c. It was actually the test of *least* similarity to Test 1. Those *Ss* who picked luck as being the factor most important in getting a high score on Test 1 tended to imitate the No. 2 man *less*, both on Test 1 and on Test 2.

In order to discover what factors might have been responsible for some of the individual differences and the inconsistencies in *Ss*' behavior, some further analyses of the data were carried out. We observed that two-thirds or more of the *Ss* in both experimental and control groups tended to choose the same answer on items 1, 2, 12, and 14. We also observed that the No. 2 men in the control and experimental groups differed in the extent to which they tended to pick these "popular" alternatives. We assumed that the tendency to give popular answers on these items reflected a *general* tendency on the part of *S* to use a simple set of hypotheses for obtaining his answers. The question arose: Would such an *S* be imitated more than one who tended *not* to pick simple hypotheses for obtaining his answers?

TABLE 3

FACTORS THOUGHT BY SUBJECTS TO BE "MOST IMPORTANT" IN GETTING A HIGH SCORE ON A TEST

TEST*	*N*	INTELLIGENCE	ABILITY TO THINK QUICKLY OR ABILITY TO CONCENTRATE	PRACTICE	LUCK
Test 1	300	5%	68%	10%	17%
Test 2a	100	10%	78%	2%	10%
Test 2b	100	1%	40%	50%	9%
Test 2c	100	2%	61%	14%	23%

* The generalization tests are arranged in order of presumed similarity to Test 1.

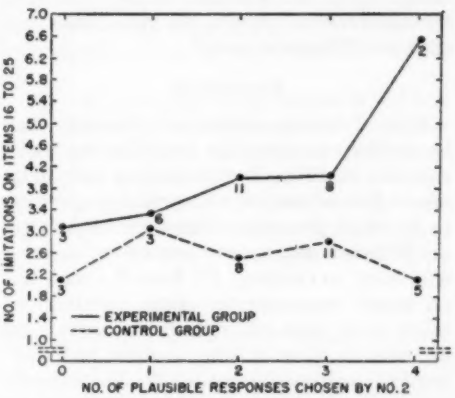

FIG. 5. IMITATION OF THE NUMBER 2 MAN IN RELATION TO THE NUMBER OF PLAUSIBLE (POPULAR) RESPONSES CHOSEN BY HIM

Each point on a curve shows the amount that the No. 3, 4, and 5 men imitated the No. 2 man on items 16 to 25 of Test 1. The small numbers below each point show the number of groups which entered into the average.

The answer to this question can be found in Fig. 5. We rated the No. 2 men according to whether they gave zero, one, two, three, or four popular answers on items 1, 2, 12, and 14, and calculated the amount that these men were imitated by the No. 3, 4, and 5 men (averaged) on items 16 to 25. It can be seen that in the experimental groups the greater the number of popular answers given by the model, the more that model was imitated. No such trend is observable in the control groups. Since the two curves never cross, a one-tailed sign test shows that they differ at the .05 level of significance.

In order to determine what effect *Ss* had on each other, an analysis was carried out for each position, relating the reactions of the man occupying that position (whether he imitated the No. 2 man, or some other man, or failed to imitate altogether) to the combination of reactions of the men in preceding positions. This analysis revealed two very definite trends: (*a*) An *S* in an experimental group was more likely to imitate the No. 2 man if someone before him had already given an imitative response on that item. If there was *any* kind of "precedent" for imitation, *S* was more likely to imitate. (*b*) If several *Ss* in the group agreed on a single answer for a given item, a later *S* was

likely to give that same answer also, even if, in the experimental groups, the No. 2 man had chosen a different answer.[9]

DISCUSSION

Some of the inconsistencies in the results can be partially explained by analyzing the social situation in which *S*s were making their judgments. Several motives were probably operating in *S*s which demanded that they *not* imitate: (*a*) Imitation might have been viewed as being equivalent to cheating. (*b*) Even if a man had no moral compunctions about cheating, he might have been afraid of being caught and punished by *E*, or of being branded by his fellow *S*s as a cheater or copycat. (*c*) The *S*s might have had a genuine desire to solve the problems for themselves, viewing them purely as an intellectual challenge rather than a test to get a high score on. (*d*) The *S*s might have felt highly competitive and reasoned that they could not beat out anyone else's score if they merely imitated.

Any or all of these motives pitted against the desire to obtain a high score by copying would produce a conflict situation which could be resolved in one of several ways: (*a*) An *S* could leave the field; this implies becoming disinterested in the whole proceedings and a relapse into aimless guessing. Only a few *S*s seemed to react in this way. (*b*) An *S* could give in to either one of the impulses and suppress the other one; this implies imitating a great deal or not at all. The number of *S*s choosing this course also seemed relatively small. (*c*) An *S* could compromise; this implies that *S* will imitate, but only part of the time. Had the tests not been multiple-choice tests, a compromise solution could have been achieved by picking answers that lay in between that of the model's and *S*'s own, but on all tests with the exception of Test 2b, an *S* had to make the decision to give either his own or the model's answer—there was no in-between alternative available. A compromise solution therefore consisted in an alternation between imitation and nonimitation. If *S*s tended to adopt such a resolution the results would reflect this in the low level of imitation and the kinds of "inconsistencies" in behavior which were previously pointed out.

It can also be argued that any of the resolutions to the conflict situation involve frustration, and that this may express itself in the form of hostility toward *E* or to the No. 2 man, who came through with a high score. If such hostility was produced, it might well have expressed itself in a desire to show up the No. 2 man, to beat him out on a new test. The test involving dot numerosity judgments offered a good opportunity to experimental *S*s to express such feelings since it was the sort of task that depended partly on practice. The *S*s might well have reasoned that they now had as good as, or a better, chance to get the correct answers than the No. 2 man. If such reactions were fairly common among *S*s, they would account for the reversal found on Test 2b.[10]

The large individual differences among *S*s can be partially accounted for by the fact that *S*s interacted with each other in a complex way. They were influenced by whether others imitated before them, by majority opinions, and by the plausibility of the answers that were given. We pointed out that No. 2 men who tended to give popular answers were imitated more than No. 2 men who did not. This suggests the possibility that *S*s imitated more if they "understood" the behavior of the model, a condition which obtained if he used simpler hypotheses to make his matches.

We may conclude by pointing out that the Miller and Dollard paradigm can be made operational for adult imitation, and that our results tended to support the theory, but that reward, defined in any terms sufficiently nar-

[9] The fact that majority judgments strongly affect an individual's judgments has been amply documented in the literature on suggestion (5, 7, 9) and has been demonstrated in an interpersonal setting by experiments such as Sherif's (8), Asch's (1), and Berenda's (3). The effects of "precedent" in condoning an imitative response have only been brought out by Asch (1, 2), however, in his study showing the effects of giving the minority members in a group some social support by having another *S* make "correct" judgments when the majority is making "incorrect" judgments. Our results support Asch's formulations interpreting *S*s' behavior as a rational adaptation to the total situation.

[10] Festinger (4) distinguishes tasks based on "physical reality" where answers can be checked against actual events from those based on "social reality" where answers can only be checked against the opinions of others. If "physical reality" is high as in the case of judgments of dot numerosity, and *S* feels he has the necessary skill to make the required judgments, social conformity may be expected to be at a minimum. Our results are in line with this expectation.

row to make experimental control possible, was not the only important variable which determined *S*s' behavior.

It is of interest to note that many *S*s actively resisted experimentally created forces toward imitation. The extent to which *S*s expressed their individuality, even while imitating, and the extent to which they attempted to avoid imitating altogether, points to an aspect of social behavior that has only recently begun to be stressed experimentally (1, 2, 3). Researches in social psychology have tended to put primary emphasis on conformity behavior and have not focussed on the ways in which members of social groups resist being influenced and find ways to express their individuality. The norms of a group not only define areas where conformity is demanded, but also define areas where deviance is permitted, in fact encouraged, if the group is to be at all creative.

SUMMARY

A study was carried out to test the Miller and Dollard theory of imitation on adult human *S*s. A significant number of *S*s "learned" to imitate a model when such imitation was rewarded, and the imitative response generalized to a similar but new situation where it was no longer rewarded. Three separate tests of generalization were run, varying in the degree of similarity to the original test situation. Positive generalization appeared on the most similar test. Some negative generalization appeared on the test of intermediate similarity, but no generalization appeared on the test of least similarity. This reversal was interpreted to be a result of the fact that *S*s perceived the test of intermediate similarity to be the least similar test.

The results tended to support the Miller and Dollard theory, but *S*s' reactions were highly complex. The results were interpreted as reflecting a complicated compromise between the various motives that were operating in the experimental situation and the degree to which the model's behavior was "understood" by *S*s.

REFERENCES

1. ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men*. Pittsburgh: Carnegie Press, 1951. Pp. 177–190.
2. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
3. BERENDA, R. W. *The influence of the group on the judgments of children*. New York: King's Crown Press, 1950.
4. FESTINGER, L., *et al. Theory and experiment in social communication*. Ann Arbor: Research Center for Group Dynamics, Univer. of Michigan, 1950.
5. MARPLE, C. H. The comparative susceptibility of three age levels to the suggestion of group versus expert opinion. *J. soc. Psychol.*, 1933, **4**, 176–186.
6. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
7. MOORE, H. T. The comparative influence of majority and expert opinion. *Amer. J. Psychol.*, 1921, **32**, 16–20.
8. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, **27**, No. 187.
9. WHEELER, D., & JORDAN, H. Change of individual opinion to accord with group opinion. *J. abnorm. soc. Psychol.*, 1929, **24**, 203–206.

Received April 13, 1953.

IS THERE A MECHANISM OF PERCEPTUAL DEFENSE? A REPLY TO POSTMAN, BRONSON, AND GROPPER

RICHARD S. LAZARUS

The Johns Hopkins University

In the April 1953 issue of this journal, Postman, Bronson, and Gropper (10) presented an argument and some data which were offered in opposition to the controversial concept of perceptual defense. Since there are some very serious flaws in their argument, and since they appear to ignore a number of relevant studies which have recently been published, it is important to offer a brief defense of the concept and a rebuttal of the position of these authors.

The Concept of Defense

The hypothetical construct of psychological defense originated, not from experimental data, but from observations of people with behavior disorders. It is useful because it offers us an explanation for the apparently illogical and contradictory behavior of persons whose reasoning, memory, and interpretation of interpersonal situations seem to deviate from a realistic evaluation of an actual state of affairs.

Perceptual defense as it is used by the personality-oriented experimenter is merely one manifestation of the general idea of a defense process which should produce observable effects in any cognitive situation where the individual's security or self-esteem is threatened, and where there is offered some opportunity for him to reduce the affective disturbance attendant upon such a threat or to gratify the motivation of seeing himself in a particular light. Therefore, some of the rules by which the process of defense operates should be the same whether the subject is asked to perceive threatening material, learn it, or recall it. Certainly these rules will be modified in terms of what we know about the specific perceptual process as compared with the learning or recall process. And the exploration of these rules from a perception or recall point of view has not even been begun systematically. Therefore, the question before us is concerned with whether or not the concept of defense, in the general sense, is reasonable.

The criticism of the data which McGinnies (9) has used to establish perceptual defense is perfectly appropriate. Actually, if the concept stood or fell on these data, it would have little support. There are two reasons why the McGinnies-type methodology is seriously defective. In the first place, when verbal stimuli which are judged emotional or threatening are placed in opposition to material which is judged nonemotional or nonthreatening, and when learning, recall, or perception of them is studied, it is absolutely impossible to establish whether differential performance in a group of subjects is associated with the emotional character of the material or other associative factors such as familiarity, attention value, etc. Yet one finds current experimenters still performing these kinds of experiments in learning, retention, and perception (2, 3, 6, 10).

Second, it was recognized long ago that the concept of defense as we now use it does not predict that all persons will deal with threat in the same way. It appears that we have not even caught up with Freud himself, who long ago decided that repression was not the only mechanism of defense. The only experiments which are relevant to the notion of a process of defense are those which look, not for main effects (e.g., mean differences in group perception of emotional vs. nonemotional material), but for interactions between different kinds of personalities and different ways of dealing with threat. The McGinnies-type experiment is therefore not an appropriate empirical relationship with which to test the notion of psychological defense. The Postman, Bronson, and Gropper attack on the concept of defense in perception on the basis of a few experiments which are not really appropriate tests of it, and without recourse to those experiments which *are* appropriate, is perhaps the height of need in perception. I shall consider later those experiments which might have been cited with better justification.

The Concept of Word Frequency

In place of the notion of defense, Postman, Bronson, and Gropper, as well as others (12), offer the notion that the differential frequency of usage, in written English, of the critical and noncritical words accounts for the perceptual

selectivity found in experiments in this area. The same argument may be used in dealing with selective recall.

If there really were a marked correlation between word frequency and recognition thresholds, if we were willing to overlook the limitations of the written-word counts that are usually used, and if we overlooked the fact that these word counts tell us little about the familiarity with the experimental words for any one individual, then this would still be a limited argument with which to attack the need and perception experiments. We should note that word familiarity is an intervening variable which is inferred from verbal behavior, just as motivation or personality is inferred. It is likely that word frequency itself is a product of the individual's personal needs and social experience. At best, it is possible to say that the correlation between needs or values and recognition thresholds may be mediated by this differential word familiarity.

But the fact is that the correlations between word frequencies and visual recognition thresholds are minimal, particularly at the higher word frequencies which are typically found in most experiments. Most of the covariation seems to occur when words of zero frequency are compared with words with slightly greater frequencies and over a very wide range. This fact may be established by examining the data reported by Solomon *et al.* (5, 13), who are the primary exponents of the word-frequency hypothesis. Little covariation seems to occur when highly frequent words of differing frequency value are used, and it is these high-frequency words which are typically employed in perceptual studies. There is not much doubt that word frequency is a variable in perceptual recognition under certain conditions, but it is not a variable of such great importance and generality as has been suggested.

It should also be pointed out that Postman, Bronson, and Gropper cite data of their own in which they assigned frequencies to their stimulus words and found that recognition of the taboo words was faster than recognition of the neutral words. Even the above authors seem to be aware of the inadequacy of the basis of the assigned word frequencies. To two words, *Kotex* and *penis*, they give the lowest frequency, matching them with such rather obscure words as *cleat* and *mixer* (even though all males have penises and therefore must be highly familiar with the concept). Words like *raped* get almost the same rating in frequency as words like *capon*, *lathe*, and *lotus*. *Whore* and *tiara* are equally matched. Now, whatever the written language looks like, it seems rather unlikely that a normal sample of college students will find such a word as *tiara* equally unfamiliar with words like *bitch* and *whore*; or *capon*, *lathe*, and *lotus* comparable with words like *raped*. Surely a vocabulary test or some other indication of familiarity with the words would serve a good purpose here.

There is also no doubt that findings in terms of mean differences in thresholds for emotional and neutral words could be made to go in either direction by the appropriate selection of words on the basis of such variables as context, contrast, sequential probabilities, and a host of other factors not directly related to word frequency. We have no way of knowing which variables were operating in this type of study. But this is irrelevant to the concept of defense.

SOME EXPERIMENTAL DATA

There are some recent data relevant to the question of psychological defense in memory and perception which might be cited here, since they are apparently not widely known. These data include differential thresholds in the perception of recall of emotional and nonemotional stimuli which have been found to correlate with clinical typologies, or with the results of additional exposure to different kinds of stresses. For example, Lazarus, Eriksen, and Fonda (7) were able to show that poor auditory perception of emotional material was characteristic of patients with hysterical personalities, while the reverse was true for obsessive-compulsives. In addition, Rosenzweig and his collaborators (11) some years ago related selective memory for successes under ego-involving conditions to impunitiveness and hypnotizability. The work of Alper (1) on recall and ego strength should also be noted in this context.

Furthermore, Eriksen (4) has shown that subjects who remembered their successes had higher recognition thresholds and slower learning for words on which they blocked in a word association test than for fast-response-time words; while subjects who showed superior recall of failures exhibited no retardation of learning, and had just about the same recogni-

tion thresholds for words on which they blocked as they had with words on which no blocking occurred.

Recently, Lazarus and Longo (8) have demonstrated that selective recall of failures in an interrupted task experiment was associated with superior recall of nonsense syllables which had been associated with a painful electric shock, while selective recall of successes was associated with better recall of nonshocked syllables. There seems to be some consistency in the way subjects deal with stress from situation to situation; that is to say, selective recall and perception depend upon personality characteristics of the subjects rather than on the specific nature of the material.

No one questions the observations that are regularly made about individuals who, in everyday life and in the clinic, misperceive and misinterpret the social events in which they participate. The extent of this perceptual or interpretive failure may range from mild "blind spots" to psychotic "lack of contact." Even if the laboratory were unable to duplicate or verify this kind of observation with its feeble experiments with words and its minimally stressful situations, the compelling force of these observations still could not be disregarded.

There is beginning to be a considerable number of appropriate experimental studies of defense mechanisms as they appear to operate in cognitive situations. Despite the age of these concepts, only a beginning has been made. Experiments have shown that it is reasonable to speak about processes of defense, at least in terms of consistent styles or habits of responding to threat. However, the experiments have certainly not gone far in identifying the rules by which the process operates. Ultimately, perhaps, experiments along these lines could tell us more specifically how defenses operate in such forms of behavior as perception and memory. But in the absence of such data, our experiments have begun to confirm the clinical belief that the postulation of such a process is both reasonable and useful. It can lead us to make predictions which would not be made in the absence of the concept. Moreover, it should be noted that a parsimonious explanation which could account for a limited series of experimental data might not be an explanation at all (much less parsimonious) of the manifold clinical observations which have led to the development of the concept in the first place.

SUMMARY

The present article is a rebuttal of the point of view expressed in this journal by Postman, Bronson, and Gropper (10), who criticized the concept of perceptual defense and argued that it is not supported by the experimental evidence. Issue was taken with these authors concerning their understanding of the concept of defense, their espousal of the notion of word frequency as an explanation of perceptual recognition, and the limited variety of studies which they criticize and cite as relevant to the defense concept.

REFERENCES

1. ALPER, THELMA G. Memory for completed and incompleted tasks as a function of personality: correlation between experimental and personality data. *J. Pers.*, 1948, **17**, 104–137.
2. COMBS, A. W., & TAYLOR, C. The effect of the perception of mild degrees of threat on performance. *J. abnorm. soc. Psychol.*, 1952, **47**, 420–424.
3. COWEN, E. L., & BEIER, C. G. The influence of "threat expectancy" on perception. *J. Pers.*, 1950, **19**, 85–94.
4. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 230–235.
5. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
6. LAFFAL, J. The learning and retention of words with association disturbances. *J. abnorm. soc. Psychol.*, 1952, **47**, 454–462.
7. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471–482.
8. LAZARUS, R. S., & LONGO, N. The consistency of psychological defenses against threat. *J. abnorm. soc. Psychol.*, 1953, **48**, 495–499.
9. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
10. POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–224.
11. ROSENZWEIG, S., & SARASON, S. An experimental study of the triadic hypothesis: reaction to frustration, ego-defense, and hypnotizability. I. Correlational approach. *Charact. & Pers.*, 1942, **11**, 1–20.
12. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.
13. SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.

Received May 18, 1953.

A THEORETICAL APPROACH TO PSYCHOLOGICAL MOVEMENT[1]

ALVIN R. HOWARD[2] AND GEORGE A. KELLY

The Ohio State University

SCIENTISTS almost universally agree that their manifold efforts are directed toward prediction and control of their subjects, animate or inanimate, human or not. In this paper, we are concerned with the human being, specifically the maladjusted individual, and the changes he evidences over a given time interval. We feel that movement is of prime interest to the student of human behavior and that a fuller comprehension of this phenomenon will enhance the effectiveness of therapeutic (broadly defined) techniques.

It is assumed that the therapeutic situation is one that is designed to hasten the process of change. We infer also that all changes in behavior follow upon the construing of these changes (2). Behavioral changes do not just happen; they are man-made. Therapy provides a framework wherein accelerated change may occur.

Saying that our ideas, our aspirations, precede the actual working out of these ideas and aspirations is another way of stating that behavioral change follows the construction of change. For example, today we plan an investigation which necessarily must precede our doing the investigation. Our ideas tend to run ahead of our "actual" performance, or our "actual" behavior. A further example, simple but cogent, is furnished in the area of literary composition. We tend to compose our thoughts faster than we can put them on paper. In the sense that we are aware of what we intend to write, we anticipate extensive movement. But if we were halted at any given point and our (actual) movement measured to that point, it would appear that little change had occurred. The writer, however, affected by his over-all construction of the situation, perceives his own movement in terms of this greater plan and thereby distorts his perceptions to that extent. More will be said of this contrasting process below.

In therapy, we offer the patient a situation in which the construction of change may be facilitated. The therapeutic situation provides him with a framework in which he may structure changes taking place. Although he may accomplish this to some extent in "unguided real life," his therapist aids him in a more rapid realization of what is transpiring. Under the therapist's guidance, he becomes more effective in perceiving the changes, extensive and limited, that he has undergone. In addition, the patient is helped to become more aware of future changes. He broadens his perspective to include more distant goals; his construction of events runs ahead of the more readily observable behavioral changes.

The reader may have noted the synonymous utilization of the terms, "movement" and "change." It is necessary that our present use of these terms be understood. In general, we regard movement or change as a constantly varying behavioral process. The individual manifests behavioral variations throughout his existence. Measurement is accomplished through arbitrary segmentation along the time dimension. Movement is here discussed in two contexts, measured and perceived. *Measured movement* refers to the variation ascribed to an individual's behavior by "outside" observers. It is the change noted in behavior deriving from observation of the individual on two separate occasions. *Perceived movement* denotes the variation ascribed by the individual to himself at two instants in time. This is a more complex process than appears at first glance. It is not simply a matter of twice reporting on himself. At one instant, his perceiving of present status will also be somewhat affected by his anticipation of future movement, and, to an even lesser degree, by constructions of prior events. At the second instant, his perception of past status is slightly affected by anticipation of future movement, and, to a much greater degree, influenced by a reorganization of the past derived through

[1] Drawn from a dissertation submitted by Howard, with the direction and assistance of Kelly, to the Department of Psychology, The Ohio State University, in partial fulfillment of the requirements for the Ph.D. degree. Grateful acknowledgement for their generous aid is made also to Drs. H. Gordon Hullfish, Francis P. Robinson, Julian B. Rotter, and Robert J. Wherry.

[2] Now at Veterans Administration Hospital, Chillicothe, Ohio.

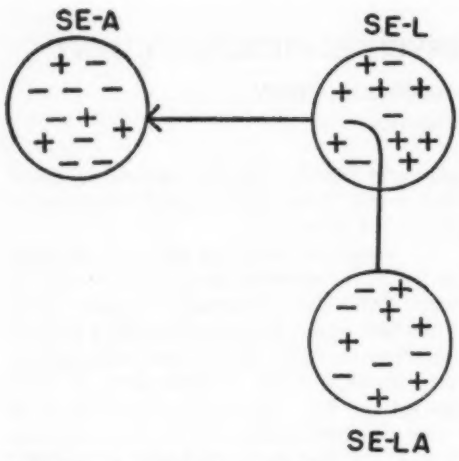

FIG. 1. OPERATION OF CONTRAST-ENHANCEMENT ON PERCEIVED STRENGTH (+) AND PERCEIVED THREAT (−)

the construction of the present. In any event, the individual's perception of the temporal changes in himself can be expected to be more distorted than that of a "neutral" observer. Examination of the effect of future anticipations upon behavior remains largely for later study. More fully treated below is the perception of past status.

It is hypothesized that perceived movement is greater than measured movement. It is our contention that this derives from a "contrast" phenomenon not unlike that which may be ascribed to the work of such investigators as Bartlett (1), Zeigarnik (4), and Ovsiankina (3). As a result of Bartlett's pioneer efforts, for example, we have since accepted the notion that perception is functionally selective. He demonstrated effectively that past experiences influence present perceptions and that recall of these perceptions changes with time through reorganization of the past. In the repeated reproduction of stories, for example, certain elements are successively distorted so that they become increasingly removed from the originals. This process of distortion we have chosen to term a contrast from the measured (or original) perception. The studies of Zeigarnik and Ovsiankina may be similarly viewed. They demonstrated the tendency of individuals to recall interrupted tasks more often than tasks which had been completed. The aim was completion, and when the aim was not met, the contrast was sufficiently great to engender greater remembrance of the failures.

It is, then, our view that upon recalling past events the individual tends to focus upon the most contrasting of these events—the sharper details out of a time telescoped; that is, an individual's segmentation of time differs from an "actual" time interval. The past becomes a comparatively more undifferentiated mass in which earlier and later events merge or telescope. From this mass the individual recalls only the "figures" or "sharper details." Concomitant with "actual" movement, the individual perceives himself to be "playing a new role(s)." As a result of these new constructions, any degree of change tends to be "overexaggerated" so that the contrast between past and present events is "overenhanced." Perceiving movement in himself, the individual will, upon contrasting what to him is the relatively static past with the ever-moving present, tend to construe a continually widening gap between his present and past behavior.

An illustration of the manner in which contrast-enhancement is deemed to operate may be helpful (Fig. 1). Let us suppose that a given individual is requested to render three self-evaluations (SE): (*a*) at instant A concerning current status at A (SE-A); (*b*) at instant L (for later) concerning current status at this subsequent point in time (SE-L); (*c*) also at instant L but concerning status at instant A (SE-LA). Let us further suppose that, when rendering SE-LA, this individual perceives himself to feel much better at instant L than he felt at an earlier point in time, A. The construction system at A, now perceived relative to the (forward) construction system at L, is viewed in "sharper contrast," i.e., the individual says, in effect: "I see this constellation at L which depicts my assets and liabilities. Now (L) I have more assets than liabilities and I know I wasn't feeling as well before this (A). In fact, according to the way I feel now, I must have been pretty low at that time." Thus, as a result of contrasting past and present, a "distorted" evaluation appears at LA. A formerly perceived threat is now recalled as a strength, and vice versa; a formerly perceived strength is now recalled as a strength of even greater intensity, etc. This apparent perceptual distortion leads one

to hypothesize that perceived movement is greater than measured movement, i.e., the difference perceived by the individual as represented by his evaluations at L and LA exceeds the difference noted by external observers at L and A. Thus L serves as a kind of buffer through which perceptions of the past are filtered and distorted. The "figures" of the pattern constituting A are heightened by this contrasting process and the "ground" becomes even further removed. But the figures are not carried over to LA unchanged. In the course of time, and as the result of filtering through L, they undergo alteration. The "outside" observer, noting behavior merely on two separate occasions does not respond to these "filtering" conditions.

A further basis for understanding the process of movement presents itself. An individual becomes aware of the variance in his construction system as events are successively construed by him. At L he construes only limited variation in that he tends to confine his perceptions to a relatively shorter period of time than he does at LA. At "point" L in the evaluation of his psychological processes, he is aware of changes which have transpired, but to a greater degree on a level of "less cognitive awareness" (2). At LA he is requested to bring this awareness to a higher (more cognitively aware) level whereupon three processes occur: (*a*) removal from restrictions of less cognitive awareness at L; (*b*) "conscious" perception of a greater number of successive constructions occurring over a longer time span tending to remove the person further from the limited perspective at L; (*c*) evaluation of the concept of change itself.

Movement may be regarded as a manifestation of the rate of organization of experience along a time line. Differences in the degree of movement between individuals may be viewed as the result of differences arising in such rates. If two persons exposed to the same (therapeutic) situation reveal significantly discrepant degrees of movement, one is obviously assimilating at a slower rate than the other. If we hold as a therapeutic objective the maximum absorption in the briefest time, we must devise means for gauging a particular individual's organization rate. As the point of diminishing returns is approached, the process of termination may be profitably initiated. If, then, we examine historical factors in order to group various "kinds" of subjects, we should be able to predict the extent of their movement in a given situation and within a specified period of time.

In this regard it is felt that case histories may indeed afford a fruitful source of hypotheses. Our predictions would cluster about the "hunch" that certain construction systems held earlier, or certain cultural characteristics to which an individual has been exposed, tend to hinder movement; others tend to facilitate movement. For example, the person who construes life apart from his mother as unbearable would be expected to anticipate backward movement in an institutional framework. On the other hand, the individual who has been actively striving for years to sever the umbilical cord may welcome a hospital environment relatively free of patronizing authority figures and construe forward movement within such a situation. We would hypothesize "regression" also on the part of adherents to an atmosphere of cultural propriety when they are exposed to the customary inelegancies of institutional existence.

We have tended to view the fundamental aim of therapy, broadly conceived, as the stimulation of more rapid change in the patient than would result from his own unguided efforts. The individual who perceives himself as remaining relatively stationary—"I don't seem to be getting anywhere"—lacks the experience of felt movement. Left to his own devices, it would appear that his future chances for rapid movement are rather slight. If therapy is to prove successful, the patient must feel that he is changing, or at least perceive himself to be capable of changing.

Extending our discussion to the classroom, we note that the teacher teaches and the learner learns. Some pupils assimilate knowledge at a faster rate than others. If a student is to realize the maximum from his attendance, he must somehow develop the notion that he is in the process of changing; for if he perceives no change, he may inquire, "What good is an education?" As he perceives that he is learning, as he perceives that school experiences are making him different from what he was earlier, his assimilation rate may be expected to accelerate (at least for a time). Perhaps test results and grades (the measured change) will

fail to reveal movement as extensive as that perceived by the student; nevertheless, if he perceives a greater change in terms of some personal norm, it should, for the serious student at least, provide sufficient incentive for further change. As long as he contrasts a prior status with his "better educated" status of the moment, he will continue to view his school experiences as desirable events.

Perhaps then, our use of objective tests (in school and elsewhere), group norms, and grades has too long overemphasized status and disregarded the essential aim, movement. More stress must be placed upon the individual's perceptions of the learning process. If he views a situation to be beneficial to him, his continuance in that situation should be facilitated. If he feels that he is learning something in the classroom, he will seek further education. Why risk stifling incentive, particularly in the formative years, with our noneradicable "final grades" which so frequently result in the inference of "final" attainment rather than in the inference of having achieved merely a given termination? Perhaps a more useful approach would be to "ask him" what he believes he has learned, and then clarify the "facts" relating to those issues on which he appears to be misinformed. In this manner, we first pay heed to the person's perceptions, and we do not hamper initiative to the degree that may occur when an arbitrary standard (a desired measured status) is imposed from above.

Viewing the phenomenon of change in terms of "contrast" permits us to link the not-too-dissimilar areas of (formal) education and therapy. In the former, we attempt to convey diverse information, often relying, in our textbooks and lectures, upon examples which tend to clarify the issues under discussion. As a result, frequently we resort to overenhancement of differences and similarities; we draw sharp contrasts, emphasizing discrepancies; and we offer oversimplified analogies, highlighting likenesses, in order to "make the point stick." Similarly, in therapy, we draw upon a relatively minor change that the patient has perceived in himself at an earlier date and utilize this to stimulate further change. We recall to him the movement he reported earlier, contrasting it with a movement he now reports, in our endeavor to encourage his anticipation of further change. As in the educational situation, we resort to overenhancement. We are relatively uninterested in the actual, or true, change that occurs. We value more highly the *perception* of change taking place than a more objective appraisal of what is "actually" occurring.

In the school this means that we foster the perception that one is learning, that one is different from what he was at a period when he knew less. Our sharp, clear-cut examples readily register with the student, with the result that he perceives movement in his rapid grasp of a concept, a solution, or a technique. In the therapeutic situation, this means that we guide the patient in the perception of contrast between earlier behaviors and contemporary ones. We strive to get him to see that he is changing or has changed. In both situations, so long as the individual perceives that he has benefited, he is less likely to terminate inopportunely. It matters little if, by all objective criteria, he has evidenced less movement than he has perceived. Indicating that he is capable of perceiving movement, he has indicated that he is capable of continuing such behavior in the future. In both situations, we help the individual adopt a point of view by means of which he considers himself to be a continuing learner; we guide him toward establishing a framework of thinking within which he may structure future changes as a learner, or as an "improved" patient. One useful principle leading toward the attainment of this goal is that of "contrast," whereby the teacher or the therapist may help the person feel that his change may well be considered more extensive than it might be in terms of some arbitrary measured standard.

We have, at several points, referred briefly to anticipations of future movement. Similarly, the sole mention of the direction movement may take has been confined to illustrative material. Simultaneous elaboration of both areas is possible.

The fundamental postulate in Kelly's "psychology of personal constructs" reads: "A person's psychological processes evolve toward what he construes to be an optimal anticipation of events" (2, p. 29). It is the "evolutionary" aspect of the postulate upon which primary stress must be placed at this point. Perhaps the person who revamps his construc-

tion system to the extent that he perceives more change in himself than does another observer (even if this is not now improvement in the customary sense) is actually in the process of anticipating greater prediction and control of his behavior in the future. As an individual initiates movement, a wider range of subsequent movement becomes potentially available to him. The more movement occurring, the more numerous the chances of eventual satisfactory "anticipation of events."

We have said that therapy is designed to hasten movement. If one "kind" of therapy is more accelerating than another, we would expect persons in the former situation to have available to them even more opportunities for anticipating events. For example, if we select two groups of hospital patients on the basis of whether this is the first confinement or a subsequent one, we would hypothesize a greater degree of movement on the part of the new admission. This would apply equally to both perceived and measured movement and derives from the theoretical deduction that first admissions are in a more accelerated therapy situation than are readmissions for two reasons. First, the new patient has not been in the same milieu before. He thus perceives more freedom to move, in whatever direction; he may either be frightened by the novelty and therefore move backward, or take advantage of the new environment and the removal from one in which he was poorly adjusted and thus change "for the better." Second, institution personnel tend to minister to the first admission patient to a greater degree than to the readmission patient. The psychiatrist conducts a more detailed initial interview with him. In rapid succession, he undergoes a series of routine examinations—physical, psychological, social service. Again he may move in either direction—backward or forward. All this attention may serve to make him feel that people are helping to cure him, or it may result in his feeling that he must indeed be very ill to have so many persons consulting him.

We have, then, what is essentially an accelerated therapy situation for the first admission. The readmission, on the other hand, has undergone the hospitalization process a number of times. For him, there is less novelty, less contrast from one confinement to the next. The routine is the same, the personnel are the same, the wards are the same. Returned to the same situation, he remains less likely to perceive changes in himself and less likely to be observed by others to have changed.

Thus we note that before an individual can change, he must construe change as being potentially available to him. Therapy may aid by structuring the need for change. Once such construction has occurred, certain behaviors will be observable to others. The person who perceives that he is moving forward will be noted to be moving forward; the person who perceives that he is moving backward will be noted to move in a backward direction. Analogous observation will be made of the person who perceives that he is at a standstill, that he is "getting nowhere." There are occasions, however, wherein an individual is observed to be changing to a greater degree than he perceives. It is felt that this trend will become reversed, in view of our earlier statements concerning behavior following one's construction system. This person will soon be observed to be behaving more in accord with his own perceptions. If he perceives backward movement, or no movement at all, his behavior will likewise reflect this. Should his perception of movement alter so that now he sees more movement in himself than others do, we have then the situation originally posited.

Thus the personal construct system is reflected through successive modifications of behavior. By periodically observing behavior, we may infer the nature of the perceptual system. Rather than content ourselves with a cessation of operations at that point, we would undoubtedly desire to gauge the extent to which a person's behavior is a consequence of his construction system. One method would be to request periodic self-evaluations of an individual. During the intervals between the rendering of these, observational records of the individual's behavior in various activities should be maintained. The primary objective would be to predict from any given self-evaluation, i.e., the construction system at a given point in time, the behavior that the individual is likely to exhibit during a specified future period. The observer's records would serve as a check on the predictions made and, of greater import, may be expected to yield principles of prediction applicable to other situations.

SUMMARY

An approach to psychological movement, or change, has been presented. Stress has been placed on the role of therapy in accelerating movement. Earlier investigations have afforded the basis for currently offering a "contrast" phenomenon to explain certain aspects of movement and to link the fields of (formal) education and therapy. Implications of these and other fields are noted.

REFERENCES

1. BARTLETT, F. C. *Remembering*. Cambridge, Eng.: Cambridge Univer. Press, 1932.
2. KELLY, G. A. The psychological construction of life. Unpublished manuscript, Psychol. Clinic Library, Ohio State Univer., 1951.
3. OVSIANKINA, M. Die Wiederaufnahme von unterbrochener Handlungen. *Psychol. Forsch.*, 1928, **11**, 302–389. (In *Psychol. Abstr.*, 1929, **3**, 2599.)
4. ZEIGARNIK, B. Das Behalten erledigter und unerledigter Handlungen. *Psychol. Forsch.*, 1927, **9**, 1–85. (In *Psychol. Abstr.*, 1927, **1**, 2365.)

Received April 14, 1953.

THE EFFECT OF EXPERIMENTAL AROUSAL OF THE AFFILIATION MOTIVE ON THEMATIC APPERCEPTION

JOHN W. ATKINSON, ROGER W. HEYNS, AND JOSEPH VEROFF[1]

University of Michigan

SHIPLEY and Veroff (9) have reported scoring criteria for thematic apperception stories to provide an index of strength of the affiliation motive (n Affiliation). Following the logic of earlier motivation studies (1, 3, 6, 7), they found that experimental arousal of n Affiliation increased the frequency of imaginative stories dealing with affective concern over separation from another person. In the present experiment, conducted for the same purpose with several differences in experimental procedure and a different set of pictures, *separation imagery*, as a criterion for designating affiliation-related stories, did not by itself discriminate between experimentally aroused and control conditions. This paper presents the experimental findings which support a broader definition of affiliation imagery and scoring procedures for obtaining a measure of strength of n Affiliation, and a reconciliation of the differences between the two studies.

The theoretical basis for the procedures employed is elaborated elsewhere (3; 5, ch. 12; 6, chs. 2, 7; 10). A series of pictures was selected to which male college students might be expected to reveal motivation for social acceptance (n Affiliation). Then, since the aim of this research is to define clearly the kinds of imaginative responses that are symptomatic of n Affiliation, two test situations were designed: one in which affiliation-motivating cues prior to the writing of stories would be maximized, and one in which such cues would be minimized. A 20-minute period of sociometric ratings conducted in the dining room of a fraternity house (also used in the Shipley-Veroff study) met the requirement for a situation containing many cues which would arouse anticipations of affective changes contingent upon being liked and accepted by others or disliked and rejected. A college classroom in which the subjects (*S*s) performed a 10-minute anagrams task following a task-orientation-type instruction fulfilled the theoretical requirement for a situation having fewer affiliation cues and, if anything, incompatible cues. It is assumed that this control condition contained fewer affiliation-motivating cues than the fraternity dining room setting without the sociometric procedure which constituted the control condition of the Shipley-Veroff experiment.

[1] We wish to acknowledge the initial support of this research by a grant from the Horace H. Rackham School of Graduate Studies, support of its continuation by a Social Science Research Council Fellowship to JWA, and the portion of RWH's time made available during 1952–53 as a result of a Ford Foundation grant for related research on conformity.

METHOD

Procedure

Aroused condition. The *S*s were 31 members of a fraternity at the University of Michigan (spring, 1951), which received $25.00 for participation in this experiment. They congregated in the dining room following an evening meal. The rationale offered for the sociometric procedures included reference to research findings on conferences which showed how important interpersonal relations were in relation to satisfaction with group performance. The sociometric procedure is described in detail elsewhere (9). It consisted of ranking traits according to the degree to which possession of the trait would make a person likeable, description of self and members of one's own class delegation in terms of these traits, and finally choice of *at least three* persons as most desirable close personal friends.

After about 20 minutes of activity of this sort (assumed to arouse n Affiliation), the imaginative measure of motivation was administered according to the procedures described in earlier work on the achievement motive (6, 7). Each *S* wrote four-minute stories in response to six pictures. The pictures used were the five originally developed by Henry and Guetzkow (4) for use in research with small groups and a sixth from the n Achievement series. The pictures in order of presentation with their code designations for the present study were: (*a*) the heads of two young men facing each other in conversation; (*b*) a man standing in an open doorway looking away from the viewer at the landscape outside; (*c*) six men seated about a table as if in conference; (*d*) an older woman with a troubled expression looking away from a younger man; (*e*) four men sitting and lounging informally in a furnished room; and (*f*) a boy seated at a desk holding his head in one hand (H of the n Achievement series).

Control condition. The 36 *S*s were male students in an introductory psychology course (spring, 1951). The *E* was introduced during a class period as a person engaged in collecting preliminary data for some other work.[2] Prior to administration of the six pictures, the

[2] We wish to acknowledge the assistance of David Klaus in running this condition.

classroom group was given an anagrams task (6). They were instructed to make as many words out of the word "generation" as they could during the limited time period. No particular emphasis was put on the importance of this task. Its purpose was to reduce the influence of any affiliation-motivating cues in the classroom situation by introducing incompatible achievement-related cues. After 10 minutes of this, the measure of motivation was introduced.

n Affiliation Scoring System

The breakdown of imaginative behavioral sequences into the categories developed earlier for hunger (1) and achievement motivation (6, 7) was accomplished for these stories in terms of affiliation. The affiliation imagery criterion, which defines the concept of n Affiliation, is given in detail. A complete scoring manual with illustrative examples is in preparation.[1]

Affiliation Imagery (Aff Im) is scored when a story contains evidence of concern, in one or more of the characters, over *establishing, maintaining, or restoring a positive affective relationship with another person(s)*. Concern of this sort is immediately evident if the relationship is described as friendship.

In the absence of a clearly stated relationship of friendliness, affiliation concern can be inferred from statements about how one person feels about another or their relationship. Statements of liking, or wanting to be liked, accepted, or forgiven, or wanting understanding or sympathy indicate the desire for an affiliative relationship.

When no statement of feeling or wanting is present, the nature of the relationship can be inferred from companionate activities such as parties, reunions, visits, or relaxed small talk as in a bull session. These are regarded as affiliative activities from which affiliative feelings can be inferred.

Often the reaction of some character in a story to separation or disruption of an interpersonal relationship is indicative of affiliative concern (9). Negative feelings such as sorrow in parting or grief over some action which has led to separation imply the desire to restore the affiliative relationship.

Unrelated Imagery (UI) is scored when there is no evidence of affiliative concern as defined above, and the remaining categories are then not scored.

The following categories are scored when the behavorial sequence has been identified as affiliation related. Each is scored only once per story.

Need (N) is scored when someone in the story expresses a desire for the establishment, maintenance, or restoration of an affiliative relationship.

Instrumental Activity is scored when there is evidence in the story that a person is doing something (including thinking) to bring about an affiliative goal state. The instrumental act is scored successful (I+), unsuccessful (I−), or of doubtful success (I?), depending on the over-all outcome of instrumental activity in the story.

Anticipatory Goal State is scored when someone in the story is anticipating goal attainment or frustration. When a person in the story thinks about the pleasantness accompanying an affiliative relationship or its restoration, the anticipatory state is scored positive (Ga+). Negative anticipations (Ga−) are indicated by thoughts about pain of rejection or future separation. Doubtful anticipations are scored Ga−.

Obstacle or Block is scored when goal-directed activity is blocked or interfered with. The obstacles may be in the environment (Bw) or within the person (Bp).

Affective State is scored when some statement indicates an affective (emotional) state associated with an affiliative activity. When someone in the story experiences joy and satisfaction in affiliation, the affective state is scored positive (G+). When pain over separation or rejection is present, the affective state is scored negative (G−).

Affiliation Thema (Aff Th) is scored when the main plot of the story has to do with establishing, maintaining, or restoring an interpersonal relationship characterized by friendship, mutual interest, and sympathetic understanding. If there is a competing behavioral sequence, this category is not scored. The thema category is an indication that affiliation-related thoughts dominated *S*'s thinking to the exclusion of other thoughts during the four-minute period in which he was writing the story.

Scoring reliability. The total scores assigned to the five story records of 20 *S*s (not previously scored) by two coders working independently correlated .96. The percentage agreement in scoring affiliation imagery (twice the number of agreements divided by the total number of times the category was scored by both coders) was 95 per cent. The percentage agreement in scoring subcategories appearing in stories already scored for affiliation imagery was 90 per cent.

RESULTS

Comparisons between the control and aroused conditions were accomplished by breaking the combined control-aroused distribution of frequencies per *S* for a particular category as near to the median as possible. Table 1 shows that the expected increases in frequency of Imagery, Successful Instrumental Activity, Positive Anticipatory Goal State, Environmental Obstacle, Positive Affective State, and Thema are significant. Four categories—Unsuccessful Instrumental Activity, Negative Anticipatory Goal State, Personal Obstacle, and Negative Affective State—failed to increase.

For the Need category, an alternative breaking point which took cognizance of the sizable number of *S*s in the control condition who showed this category only once yielded clearer evidence of a difference than did the approximate median break (Table 1). While recognizing that a comparison suggested by examination of the data favors acceptance of a chance difference, we nevertheless accept

[1] Prior to publication, early drafts may be obtained from the authors.

TABLE 1

NUMBER OF *Ss* IN CONTROL AND AROUSED CONDITIONS ABOVE AND BELOW BREAKING POINT OF COMBINED CONTROL-AROUSED DISTRIBUTION OF FREQUENCIES FOR VARIOUS AFFILIATION-RELATED IMAGINATIVE CATEGORIES

IMAGINATIVE CATEGORY	BREAKING PT. (Range 0–6)	CONTROL ($N = 36$) ABOVE	CONTROL ($N = 36$) BELOW	AROUSED ($N = 31$) ABOVE	AROUSED ($N = 31$) BELOW	χ^2	p*
Affiliation Imagery (Im)	2–3	11	25	23	8	12.69	.001
Need (N)	0–1	21	15	22	9	1.16	.15
	(1–2)**	(5)	(31)	(11)	(20)	(4.27)	(.025)
Instrumental Activity							
Successful (I+)	1–2	7	29	12	19	3.04	.05
Doubtful or unsuccessful (I−)	0–1	11	25	12	19	.49	
Anticipatory Goal State							
Positive (Ga+)	0–1	5	31	12	19	5.42	.01
Negative (Ga−)	0–1	11	25	12	19	.49	
Obstacle							
Personal (Bp)	0–1	11	25	8	23	.18	
Environmental (Bw)	1–2	6	30	13	18	5.23	.025
Affective State							
Positive (G+)	0–1	15	21	20	11	3.49	.05
Negative (G−)	0–1	22	14	18	13	.64	
Thema (Th)	1–2	9	27	15	16	3.96	.025
n Affiliation score (based on Im, N, I+, Ga+, Bw, G+, Th)	7–8	10	26	21	10	10.70	.001

* In the predicted direction.

** Figures in parentheses indicate alternative breaking point, not the approximate median.

this category as symptomatic of n Affiliation pending further experimentation.

An over-all index of strength of the affiliation motive in each *S*, the *n Affiliation score*, was computed by assigning a value of +1 for each occurrence in an *S*'s stories of the seven categories of response which discriminated between presumed low and high motivation conditions. The median n Affiliation scores of the control and aroused conditions were 6.14 (range 1–15) and 9.75 (range 2–18), respectively. The expected difference between the two conditions is significant at the .001 level (see Table 1).

DISCUSSION

Shipley and Veroff (9) defined affiliation imagery exclusively in terms of separation anxiety, i.e., affective concern over separation from another person or persons. The present results argue for a broader definition of affiliation imagery and hence a broader conception of the affiliation motive. Any evidence of concern over establishing or maintaining a positive affective relationship with another person(s) (in addition to concern over restoring broken relationships) is viewed as symptomatic of motivation to affiliate according to the present results.

The difference between the two scoring systems can be attributed to (*a*) differences in the pictures used to elicit stories in the two studies, and (*b*) differences in the presumed low motivation control conditions.

Effect of picture differences. Five pictures used by Shipley and Veroff were selected from the TAT, noted for its ability to arouse fantasies concerning basic interpersonal problems. A fifth picture was especially designed by them to elicit thoughts concerning separation from others. The present pictures, on the other hand, portray a greater variety of common everyday situations. From this we expect and find important differences in the kinds of stories elicited. When the Shipley-Veroff stories were rescored according to the present criteria, comparison of the frequencies of various imaginative categories in the almost identical aroused conditions of the two experiments showed that the earlier pictures elicited disproportionately more Needs, Obstacles, and Negative Affective States than the present pictures.[4] The pictures of the earlier investiga-

[4] Tables showing the frequency of imaginative responses to different pictures in the two studies have been deposited with the American Documentation Institute. Order document 4095 from the ADI Auxiliary Publications Project, Photoduplication Service, Library

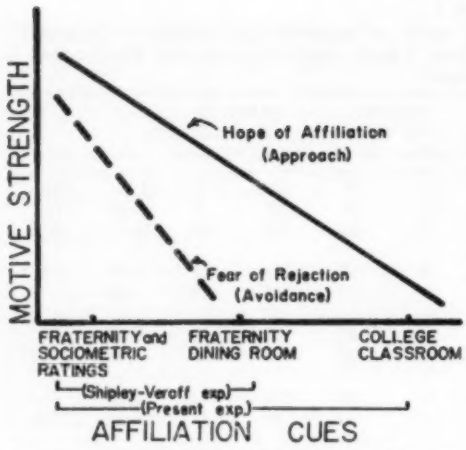

FIG. 1. ASSUMED MOTIVATING EFFECT OF CUES IN THE AROUSED AND CONTROL CONDITIONS OF THE TWO EXPERIMENTS

tion, in other words, were structured to favor perception of them as affiliation-deprivation situations. This accounts in part for the relatively restricted definition of affiliation imagery growing out of the earlier experiment.

As might be expected, the more restricted scoring procedures did discriminate between the control and aroused conditions of the present experiment, but less adequately than the more comprehensive criteria proposed here. The median score obtained for the control condition of the present experiment was 3.86 and for the aroused condition 5.00 ($\chi^2 = 3.23$, $p = .05$ in the predicted direction).[5]

To expect, in addition, that the broader scoring system proposed here would discriminate between the two conditions of the earlier experiment as adequately as the earlier scoring system would ignore the effect of an important difference between the control conditions of the two experiments on the scoring procedures that were developed by comparing the particular control and aroused groups in each experiment.

Effect of differences between control conditions. Figure 1 contains our conceptualization of the motivating effects of the cues in the sociometric arousal and control conditions of the two experiments. The general logic of the usual diagrammatic presentation of approach-avoidance conflict is implied in this conceptualization (3, 8).

The approach gradient (positive motivation to affiliate, or hope of affiliation) is presumed to be at its maximum as a result of the combined effect of cues in the fraternity dining room settings and the sociometric rating procedures in the nearly identical aroused conditions of the two experiments. The Shipley-Veroff control condition, a fraternity dining room after supper and including a food preference inventory before the imaginative measure, is assumed to have contained more cues which would arouse motivation to affiliate than the more neutral college classroom condition of this experiment which included some incompatible achievement-motivating cues.

A fear of rejection or avoidance gradient is also plotted. It represents the degree to which anticipations of painful rejection are aroused by the situational cues. It is strongest in the aroused conditions of both experiments and falls off sharply since it is assumed to be primarily determined by the sociometric rating procedure which virtually guaranteed a certain amount of rejection for all *S*s.

In the present study, the arousal of both hope of affiliation (approach) and fear of rejection (avoidance) is low in the control condition and high in the aroused condition. In the earlier study, the difference between control and aroused conditions was relatively greater for fear of rejection than for hope of affiliation.[6] As a result, the Shipley-Veroff scoring system based on pictures already favoring affiliation-deprivation imagery should be sensitive to the expression of affiliation motivation

of Congress, Washington 25, D. C. remitting in advance $1.25 for 35 mm. microfilm or $1.25 for photostats. Make checks payable to Chief, Photoduplication Service, Library of Congress.

[5] The earlier report (9) recommends scoring +1 for Affiliation Imagery, Instrumental Activity (+, −, or ?), Affective States (positive or negative), Obstacles (personal or environmental), and Thema; and −1 for Unrelated Imagery. In this comparison, Unrelated Imagery was scored 0 instead of −1 since the frequency of Unrelated Imagery is completely dependent upon the frequency of Affiliation Imagery, and scoring −1 is gratuitous, equivalent to weighting the imagery category twice as heavily as any other category.

[6] In light of the positive relationship between approval seeking and interest in eating reported by Blum and Miller (2) in work on the "oral character," one might argue that the food preference inventory included as part of the control procedure may have provided additional cues associated with affiliative tendencies.

when rejection is threatened, but rather insensitive to the effect of an increase in positive motivation alone because their control condition was already, to a certain extent, positively motivated.

Since the present control group was not influenced by either the positive-motivating cues of the fraternity dining room or the fear-producing cues of the sociometric arousal procedure, the present scoring criteria should be sensitive to positive motivation aroused by the fraternity dining room cues in the Shipley-Veroff control condition. This expectation is confirmed. Our scoring of the Shipley-Veroff aroused condition for affiliation imagery was virtually identical with the original scoring of it. However, where they report a mean (per *S*) imagery count of only 2.53 in their control condition, the more comprehensive imagery criteria suggested here produce a higher mean of 2.93. As a result, the total number of subcategories scored in their control stories is also increased, and the over-all difference between the control and aroused conditions is found to be much less than when separation imagery alone was the basis for comparison.

The scoring procedures developed in the present experiment are recommended as a more general definition of affiliation motivation than those developed in the earlier study in which the particular properties of the pictures and the nature of the difference between control and aroused conditions may be assumed to have had the effect of restricting the definition of affiliation imagery to separation anxiety as outlined in the arguments above.

n Affiliation and popularity. There is one bit of evidence to indicate that both scoring procedures were measuring the same variable in the aroused conditions of the two studies. During the sociometric rating procedure of both studies, *S*s were asked to rank a number of traits in order of the degree to which possession of that trait would make a person likeable, and later they were to assign two of the traits on the list to each member of their delegation. The mean rank of traits assigned to a person is an index of the degree to which he is considered likeable by his associates. In the Shipley-Veroff study, the rank correlation (tau) between individual n Affiliation scores and this popularity index was $-.33$ ($p = .05$). In the present study, the rank correlation (tau) between these same measures was $-.27$ ($p = .05$) and rose to $-.36$ when the one case having the lowest n Affiliation score and the least likeable ratings was eliminated.

n Affiliation is an approach tendency. Analysis of the traits used to describe *S*s during the sociometric rating procedure supports a conception of the affiliation motive as a disposition to move toward others to elicit positive affective responses from them. The *S*s above the median n Affiliation score in the aroused condition were described as *approval seeking* by at least one other person significantly more frequently than *S*s below the median n Affiliation score ($\chi^2 = 7.43$, $p = .01$). They were, in addition, described as *self-assertive and confident* (indicative of positive anticipations) by at least one other person more frequently than *S*s low in n Affiliation ($\chi^2 = 2.88$, $p = .10$). Unanticipated by us was the additional finding that they were also more often rated by at least one other person as egotistical than were *S*s low in n Affiliation ($\chi^2 = 7.43$, $p = .01$).

The correspondence of these results with the Blum and Miller (2) findings on children that indices of orality are positively related to approval seeking and negatively related to sociometric status provides a promising lead concerning the origins of the affiliation motive.

SUMMARY

The purpose of this study was to develop a method of scoring thematic apperception stories to measure strength of motivation for social acceptance, or n Affiliation. Imaginative stories were written in response to pictures by two groups of male *S*s under experimental conditions designed to differ in the degree to which motivation to be accepted and liked by others would be aroused. The presumed high motivation condition consisted of a sociometric procedure conducted in a fraternity dining room before stories were written. The low motivation control condition was a college classroom.

Behavioral sequences dealing with attempts to establish, maintain, or restore positive affective relationships with other persons occurred more frequently in the imaginative stories of the group in which n Affiliation had been aroused. The differences between the scoring definitions proposed here and a more restricted set growing out of a similar experiment by

Shipley and Veroff are reconciled in terms of an analysis of differences between the pictures and low motivation control conditions in the two experiments.

REFERENCES

1. ATKINSON, J. W., & MCCLELLAND, D. C. The projective expression of needs. II. The effect of different intensities of the hunger drive on thematic apperception. *J. exp. Psychol.*, 1948, **38**, 643–658.
2. BLUM, G. S., & MILLER, D. R. Exploring the psychoanalytic theory of the oral character. *J. Pers.*, 1952, **20**, 287–304.
3. CLARK, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.
4. HENRY, W. E., & GUETZKOW, H. Group projection sketches for the study of small groups. *J. soc. Psychol.*, 1951, **33**, 77–102.
5. MCCLELLAND, D. C. *Personality*. New York: William Sloane Associates, 1951.
6. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
7. MCCLELLAND, D. C., CLARK, R. A., ROBY, T. B., & ATKINSON, J. W. The projective expression of needs. IV. The effect of need for achievement on thematic apperception. *J. exp. Psychol.*, 1949, **39**, 242–255.
8. MILLER, N. Comments on theoretical models illustrated by the development of a theory of conflict behavior. *J. Pers.*, 1951, **20**, 82–100.
9. SHIPLEY, T. E., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.
10. VEROFF, J., WILCOX, S., & ATKINSON, J. W. The achievement motive in high school and college age women. *J. abnorm. soc. Psychol.*, 1953, **48**, 108–119.

Received May 2, 1953.

INTERPERSONAL KNOWLEDGE AND INDIVIDUAL AND GROUP EFFECTIVENESS

F. LOYAL GREER

Institute for Research in Human Relations

EUGENE H. GALANTER AND PETER G. NORDLIE

University of Pennsylvania *University of Maryland*

THE major hypothesis tested in this research was that the knowledge of interpersonal relationships by individuals in a group is related to effective group behavior. The more one knows about his physical world, other things equal, the better able he is to move in that world to achieve his purposes. It seems reasonable to extend this concept to the social environment.

A study by Chowdhry and Newcomb (2) provides evidence which indicates that sociometrically nominated leaders of face-to-face groups have more accurate perceptions of their groups' attitudes toward group-relevant issues than nonleaders and isolates. Such leadership can be conceived as indicative of personal effectiveness in the group situation. The hypothesis follows that groups composed of individuals with more accurate social perceptions about their group than groups with less perceptually accurate individuals will be more productive and effective.

The problem was to locate a criterion which would provide a means of differentiating between groups in terms of their effectiveness or productivity, and in situations where the effectiveness of the group would depend on the teamwork of its members. In order to examine further the finding concerning the accuracy of social perceptions of the leader as compared with followers, the groups needed a leader-follower structure. In addition, these groups had to have past histories as groups. Otherwise, differences would have been unlikely between leaders and followers or between groups in the accuracy of the perceptions of the interpersonal relationships and attitudes of the group.

METHOD

Criterion. The above requirements for the research were met in a study conducted by the Institute for Research in Human Relations under contract with the Personnel Research Section of The Adjutant General's Office. Havron, McGrath, and Fay (4) developed a criterion performance problem designed to measure the effectiveness of infantry rifle squads in carrying out common combat missions. This criterion was then used to validate various predictors. Moreover, the scores on the criterion field problem for infantry rifle squads made it possible to test the hypotheses in this study.

The criterion field problem was based on Army field manuals and consultations with military tacticians at The Infantry School, Fort Benning, Georgia. Regardless of the ultimate validity of the criterion, the scores obtained on the problem depended on the teamwork of the nine men in a squad. The criterion was a six-hour squad field problem. There were four different missions: (*a*) the squad as a point of advance guard, (*b*) the squad as a reconnaissance patrol, (*c*) the squad in a defense position, and (*d*) the squad as an attacking unit. These performances were objectively scored by trained observers in terms of very specific actions. Interrater reliability was +.88 (4).

The Army infantry rifle squad includes a squad leader, an assistant squad leader, an automatic rifleman, an assistant automatic rifleman, and five riflemen. These men must work together as a team in the field in order to achieve group goals. The men used in this study had lived and worked together anywhere from a month to two years.

Hypotheses. Three hypotheses were formulated:

1. Squad leaders will be more accurate in their perceptions of the preference hierarchy structure within their squads than will other squad members.

2. The more popular individuals in a squad will have more accurate perceptions of the preference hierarchy structure in their squads than will the less popular men.

3. Effective groups as measured by the criterion field problem will be characterized by members with greater accuracy in the perception of the preference hierarchy structure in their squads than by members in less effective groups.

Procedure. The measures of the group-relevant attitudes and perceptions were obtained during individual interviews. A squad member was asked first to list the men in his squad according to how well he liked them. He then was asked to list them in the order that he thought they were liked by the group as a whole. In this second listing the interviewee included himself. The first listing was used to determine the ranking of a man in the preference hierarchy structure of his group. Each individual in a squad was ranked in this preference hierarchy structure according to the average listing position given to him by all of the other men in his squad. This derived preference hierarchy ranking was compared with each squad member's perception of the preference hierarchy structure in his squad. From this comparison a discrepancy score was obtained for each

man. The larger the discrepancy score, the less accurate the interviewee's perception of the preference hierarchy structure within his squad was assumed to be.

Subjects and interview. Sixty-three infantry rifle squads were tested on the criterion field problem at Camp Atterbury, Indiana. The members of the 13 highest-scoring or effective squads and the members of the 13 lowest-scoring or ineffective squads were interviewed by eight different civilian interviewers. Neither interviewers nor interviewees knew whether a squad member came from an effective or an ineffective squad. Assurances were given to squad members that what they said would not be revealed to anybody in the Army. Toward the end of the interview the questions relevant to this research were asked.

Results

Leaders vs. nonleaders. Discrepancy scores for the accuracy of the perception of the preference hierarchy structure within a group could be as high as 40. The actual range for this sample was between 4 and 35. The 23 squad leaders interviewed had a mean discrepancy score of 15.0. The inability to interview 18 squad members did not bias any of the results. It meant only that the attitudes of eight men instead of nine determined the preference hierarchy ranking for some squads. The mean discrepancy score for the 193 nonleader men interviewed was 16.5. The difference between distributions was not statistically significant at the .05 level using chi square as the test. The obtained chi square was 2.70 with 1 *df*. The required value for chi square is 2.706 for significance at the .05 level using the one-tailed test. However, another analysis was made to check this finding. Mean discrepancy scores were determined for men in each of the nine positions in a squad. The expectation that the mean discrepancy score for squad leaders would be lower than the mean discrepancy scores for men in the other eight positions was confirmed. With the use of the standard error of the frequency this result produced a t of 2.48 ($p < .05$). On the basis of these two analyses the data appear to corroborate the findings of Chowdhry and Newcomb (2). The appointed leaders were significantly more accurate than nonleaders in the group-relevant social perception.

Popular vs. less popular. The derived preference hierarchy ranking for each man was used as the first index of individual popularity. Men in ranks one through four were combined and called "popular." Men in ranks five through nine were termed "less popular." The mean discrepancy score for the 97 popular men was 15.6; for the 119 less popular men it was 17.3. The difference between distributions produced a chi square of 3.15 with 1 *df*, which is significant beyond the .05 level for the one-tailed test.

An independent and somewhat different measure of popularity was available. Included in a battery of instruments administered to the men before they went on the field problem was a 10-item sociometric test (4). Each man responded to the items in this test according to whether he would like to do a particular activity with each of the other eight men in his squad. These data yielded a figure representing the summation of positive and negative choices received by each man. The distribution of sociometric scores was split in half; those with a high sociometric score were called the "sociometrically popular" individuals, and the low scorers were called the "sociometrically less popular." The sociometrically popular individuals were found to have the more accurate social perceptions. A significant chi square was obtained for the difference between distributions with a value of 4.05 with 1 *df* ($p < .05$).

Since 17 of the 23 squad leaders were in the sociometrically popular group, it seemed that the results might have been confounded. Even though the squad leaders were eliminated in another analysis, the sociometrically popular group retained its lower mean discrepancy score. The chi square for the difference between the distributions was 4.56 with 1 *df* ($p < .05$). Popularity, as measured in two different ways, had a positive relationship with the accuracy of perceptions of interpersonal attitudes; the more popular men perceived the social relationships more accurately than did the less popular men.

Effective vs. ineffective groups. Leaders and popular squad members, in a sense the effective members in a squad, had more accurate perceptions of the preference hierarchy structures in their squads. The hypothesis follows that groups composed of men with more accurate perceptions of the social relationships in their groups will be more effective or productive groups. The 107 members of effective squads had a mean discrepancy score of 15.8. The 109 people from the ineffective squads had a mean discrepancy score of 17.1. The chi square for the difference between the distributions is

significant beyond the .05 level with a value of 4.24 with 1 *df*.

The squad leaders of effective squads had a mean discrepancy score of 13.5; for nonleader men of effective squads it was 16.1. The squad leaders of ineffective squads had a mean discrepancy score of 16.4; for nonleaders of ineffective squads it was 17.2. These data might seem to indicate that the difference between effective and ineffective squads in terms of their perceptual accuracy was essentially due to the especially good ability of the squad leaders of effective squads. But the comparison of the distributions of scores between the members of effective and ineffective squads after squad leaders were excluded produced a chi square of 3.06 with 1 *df*. This is significant beyond .05 level using the one-tailed test. Therefore, the difference originally obtained between all men in effective and ineffective squads was still significant, even though the scores of squad leaders were excluded from the analysis.

Other analyses. The possibility existed that the measure of perceptual accuracy used in this study might have been associated with some other variable which was causally related to the effectiveness of squads on the criterion field problem. The variables of age, education, Army grade, or whether a person came from a small or large community did not differentiate between men in effective and ineffective squads. The length of time the men had worked together might have affected both the accuracy of social perceptions and the effectiveness of the squads. The data indicated that members of effective squads had an average of 168.1 days together before the field problem, while members of ineffective squads had an average of 170.4. The difference is negligible and inconsistent with what might have been expected.

An Army measure of intelligence showed no significant difference between the mean scores of all of the members of effective and ineffective squads. There was a significant difference, though, between squad leaders; the squad leaders of effective squads had a significantly higher mean intelligence score than squad leaders of ineffective squads (5). The relationship between the intelligence of a squad leader and the effectiveness of his squad is not surprising. Since intelligence and the accuracy of social perceptions may be related, one might wonder whether the intelligence of squad leaders alone, and not the accuracy of social perceptions, was the causal variable related to group effectiveness. But the finding that nonleader men in effective squads had significantly lower discrepancy scores than the nonleader men in ineffective squads, but did not differ in intelligence, militates against this explanation. Both the intelligence of squad leaders and the accuracy of the social perceptions of the squad members appear to have contributed to squad effectiveness.

DISCUSSION

Hites and Campbell (6) found that leaders, both those appointed and those nominated, were no better than nonleaders in the accuracy of their perceptions of attitudes held by the members of their groups. Newcomb[1] has proposed that for certain attitudes, e.g., those attitudes within the group which everybody has had an ample opportunity to learn, there will be no difference between leaders and nonleaders in the perceptual accuracy of their estimates. There are certain other types of attitudes regarding group morale and leader competency for which the leader's perception may be distorted through his ego-involvement. In the Chowdhry and Newcomb study (2) it was found that leaders were more accurate than nonleaders in perceiving "consensus data" when relevant group issues were involved; no significant difference existed between leaders and nonleaders when nonrelevant group issues were concerned. Naturally, when all the members of a group hold the same attitude, a difference in the estimate of the group attitude between leaders and nonleaders can hardly be expected. Indeed, accurate perceptions of group attitudes may sometimes occur simply because one holds the same attitude as most of the members of his group. The negative findings of Hites and Campbell (6) can probably be explained as a function of the attitudinal issues used in their study.

Bell (1) found that in small, task-oriented, face-to-face groups with no previous association, the sociometrically determined leaders did not have greater accuracy than other group members in the perception of the attitudinal structures in the group. He explained his negative findings in terms of the lack of previous associations among members in his

[1] Newcomb, T. M. Personal communication. December 16, 1952.

groups and the fact that leaders were chosen merely on the basis of their ability to meet the group goals and not the individual needs of the group members. The positive findings in the study reported in this paper can probably be attributed to the fact that none of the conditions which appear to have led to the negative or indecisive results in the studies of Hites and Campbell (6) and Bell (1) was present.

CONCLUSIONS

Leaders and popular individuals as well as members of effective groups had greater knowledge of the group preference hierarchy structure of members within their groups. It seems reasonable to assume that accuracy in this particular perceptual category reflects accuracy in other perceptual matters relating to group attitudes toward group-relevant issues. The position of an appointed leader within a group appears to facilitate the learning of group attitudes. A popular individual also achieves, or because of his popularity comes to have, this more accurate social perception. The results support the work of Chowdhry and Newcomb (2) and indicate, furthermore, that the accuracy of the group members' perceptions of group attitudes toward group-relevant issues is related to the effectiveness of group performance.

It is suggested that this ability to perceive accurately the attitudes of another has the following consequences. If one can anticipate the needs and behaviors of other people, he can act to satisfy them. By satisfying the needs of others, an individual becomes in effect a "problem solver" for the other person. Research (3) indicates that such problem solvers are often chosen as leaders; the more a leader is perceived as a problem solver, the more the followers appear to be motivated to help the leader (5). A person possessing greater accuracy in social perceptions can act with more certainty and confidence in the consequences of his interpersonal behavior. He is in a position not only to achieve with more certainty the goals of others, but also the social goals that he has for himself.

When one group has individuals in it who are better than individuals in another group in anticipating the needs and behaviors of their group members, this group may be a more effective group in its performance. The individuals may have to be less motivated toward behaving in relation to each other than behaving in relation to the group goal. Further study in this area seems relevant for the development of a theoretical structure for the analysis of interpersonal relationships and group behavior.

SUMMARY

1. Measures of discrepancy between an objective group preference hierarchy structure of members and an individual estimate of this preference hierarchy structure were obtained for members of infantry rifle squads who had scored high and low on a criterion field problem.

2. The appointed leaders of these ongoing groups were better able to estimate the preference hierarchy structure than were nonleaders.

3. Popular individuals were more accurate in this social perception than were less popular individuals.

4. Members of effective groups were more accurate in their estimates of the group preference hierarchy structure than were members of ineffective groups.

5. The findings are interpreted and explanations are offered for reconciling the conflicting data found in other studies.

REFERENCES

1. BELL, G. The relationship between leadership and empathy. Unpublished doctor's dissertation, Northwestern Univer., 1952.
2. CHOWDHRY, K., & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
3. GREER, F. L. Neighborhood leaders. In D. Courtney *et al.*, *Naval, neighborhood, and national leadership*. Series 1953. Institute Report #1. Philadelphia: Institute for Research in Human Relations, 1953.
4. U. S. DEPT. ARMY. ADJUTANT GENERAL'S OFFICE. PERSONNEL RESEARCH SECTION. The effectiveness of small military units. Parts I–III. *Personnel Res. Sect. Rep.*, 1952, 980. 150 p. Washington: American Documentation Institute, Library of Congress, Document No. 3969, microfilm, $5.75; photocopy, $18.75.
5. U. S. DEPT. ARMY. ADJUTANT GENERAL'S OFFICE. PERSONNEL RESEARCH SECTION. An interview study of human relationships in effective infantry rifle squads. *Personnel Res. Sect. Rep.*, 1952, 983. 183 p. Washington: American Documentation Institute, Library of Congress, Document No. 3965, microfilm, $6.75; photocopy, $23.75.
6. HITES, R., & CAMPBELL, D. T. A test of the ability of fraternity leaders to estimate group opinion. *J. soc. Psychol.*, 1950, **32**, 95–100.

Received May 4, 1953.

IRREGULARITIES IN JUDGMENT DATA COLLECTED BY THE METHOD OF EQUAL-APPEARING INTERVALS[1]

SAM C. WEBB

Emory University

A NUMBER of investigators have noted various types of irregularities in sortings of attitude items made by the method of equal-appearing intervals. Types of irregularities most frequently noted have been reversal of ends of the continuum (1, 3, 5), failure to rate items (11), bunching of items in extreme intervals (2, 3, 4, 5, 10, 11), failure to use some intervals (8, 11), and misunderstanding of, or failure to, follow directions (1, 2, 6, 9, 10, 11). Though these types of irregularities have been noted, no systematic study of the extent to which they appear in judgmental data, or of their effects on scale and ambiguity values, has been made.

The purpose of this note is to report an analysis of the various types of irregularities found in the sortings made by 712 college students of the 130 Thurstone-Chave (10) statements of attitude toward the church. Hypotheses relative to the causes of some of these irregularities are suggested.

METHOD

Procedure. The method of collecting the judgments was a modification of the Seashore-Hevner technique (7). Ratings using 11, 7, and 5 intervals were recorded on specially prepared IBM answer sheets which provided appropriate numbers of identified intervals.

In preparing the booklet of items, the first three items were selected so that the first and third items would express extremely unfavorable attitudes, while the second would express a very favorable attitude. The remaining 127 items were arranged in scrambled order. To minimize reversals, the verbal designations associated with the extreme and middle intervals of the scale were written at the top of each page of the booklet.

The original Thurstone and Chave directions (10), modified to provide for recording ratings on the answer sheet, were used. Ratings were made by groups of 20 to 30 students during a 50-minute class period. The directions were read aloud to the students, and all questions concerning them were answered before the rating of the items started.

After the judges started rating the items, the administrator circulated among them to check their rating procedures. If the pattern of a judge's ratings was not such that the first and third statements fell on the "depreciation" side of "neutral" and the rating of the second statement fell on the "appreciation" side of "neutral," the administrator asked the student to explain his judgment for the item or items which did not conform to this pattern. While some students indicated that they did not fully understand what to do, a surprising number replied, "but I don't agree with it!" From the emotional tone of these replies, the writer inferred a personal involvement with the item on the part of the student and not just a confusion in directions.

Subjects. The 11-interval data were collected from 243 students enrolled in psychology classes at Georgia Institute of Technology. The seven-interval data were collected from 244 students enrolled in psychology classes at the University of Georgia; and the five-interval data were collected from 225 students enrolled in psychology and sociology classes at Emory University.

RESULTS

After a preliminary examination of the sortings for types of irregularities, nine classifications or categories were established. These were arranged in order of the seriousness of the defect as viewed by the writer. The most serious defect was placed at the top of the list. Each set of ratings was then considered for inclusion in each category, working from top to bottom. If a set was classified in one of the top categories, it was not examined for other irregularities which would permit it to be classified in a lower category. The classifications and the percentage of sets of ratings for the 11-, 7-, and 5-interval scales and for all scales combined found in each classification are shown in Table 1. "Double marking" refers to two ratings for an item. Wicker's (11) extrapolation of the Thurstone and Chave criterion was used to determine if too many items had been judged to fall in one interval.

An inversion is defined as the rating of an item on one side of "neutral" when it would be expected to fall on the opposite side of "neutral." The number of inversions per set of ratings was determined in the following way. The writer by inspection selected 49 extreme items (23 favorable and 26 unfavorable) which were fairly evenly distributed throughout the

[1] This study has been supported by a grant from the Emory University Research Committee. The writer wishes to express appreciation to Drs. Joseph E. Moore, Georgia Institute of Technology, Florene Young, University of Georgia, and to the staffs of the psychology and sociology departments of Emory University for assistance in securing subjects for this study.

TABLE 1
SUMMARY OF IRREGULARITIES FOR JUDGING GROUPS

CATEGORY	11 INTERVAL		7 INTERVAL		5 INTERVAL		TOTAL	
	N	%	*N*	%	*N*	%	*N*	%
1. Incomplete	8	3.3	9	3.7	2	.9	19	2.7
2. Double marking	16	6.6	15	6.1	2	.9	33	4.6
3. Inversions (10 or more)	8	3.3	11	4.5	10	4.5	29	4.1
4. Omissions (2 or more)	7	2.9	7	2.9	0	0	14	2.0
5. Too many in one interval	30	12.3	18	7.4	0	0	48	6.7
6. Unused intervals	9	3.7	3	1.2	3	1.3	15	2.1
7. One omission	24	9.9	16	6.6	11	4.9	51	7.2
8. Inversions (less than 10)	70	28.8	111	45.5	91	40.4	272	38.2
9. Perfect sorting	71	29.2	54	22.1	106	47.1	231	32.4
Total	243	100.0	244	100.0	225	99.9	712	100.0

test booklet. All sets of ratings were "scored" on these items to determine the number of items placed on the unexpected side of the "neutral" interval. Sortings with 10 or more inversions were placed in category 3; those with less than 10 inversions were placed in category 8.

It seems plausible to classify the eight categories which describe the irregularities into four groups. These groups and their percentage of irregularities are as follows: (*a*) Irregularities resulting from a time limitation (category 1), 2.7 per cent. (*b*) Irregularities associated with the answer sheet method of collecting the data—double markings and omissions (categories 2, 4, and 7), 13.8 per cent. (*c*) Irregularities usually attributed to "carelessness" in sorting—too many in one interval and unused intervals (categories 5 and 6), 8.8 per cent. (*d*) Irregularities defined as inversions (categories 3 and 8), 42.8 per cent.

There were fewer irregularities in the 5-interval data than in the 11- and 7-interval data. For the five-interval data 47.1 per cent of the ratings were classified as perfect, while only 29.2 per cent and 22.1 per cent of the 11- and 7-interval ratings, respectively, were so classified.

Additional analyses of inversions were made. First, a count of the number of inversions found in each of the 301 sets of ratings which were classified in category 3 (10 or more inversions) and category 8 (less than 10 inversions) showed that 127 (41.5 per cent) of the sets of ratings had only one inversion and that 259 (86 per cent) did not contain more than five inversions. Second, using all sortings in categories 3 through 9, the percentage of inversions of each of the 49 items was obtained. The distribution of these percentages showed a significant shift toward increased percentages of inversions when compared with the distributions of the percentages of inversions for the same items reported by Thurstone and Chave (10) and Wicker (11), both of whom also excluded judges before tallying judgments. However, when the percentages of inversions were computed on the basis of all sets of ratings in categories 4 through 9, the distribution of percentages so obtained was similar to the distributions of percentages for the Thurstone and Chave and for the Wicker data.

For the 49 items the percentage of inversions computed on the basis of sortings in categories 3 through 9 and the percentages computed on the basis of ratings in categories 4 through 9 correlated .63 and .67, respectively, with the percentage of inversions found in the Thurstone and Chave data for these items.

For the sortings of the 29 judges with 10 or more inversions (category 3), a detailed analysis was made by tabulating on a sheet, with statements as rows and with judges as columns, each item inverted by each judge. Since the inversions for each judge were fairly well scattered among the entire set of 49 items, they do not appear to be a function of systematic reversal of the ends of the continuum. For each person the number of inversions of favorable and unfavorable items was determined; and for each judge the obtained frequencies were tested against a 23:26 distribution (the number of favorable and unfavorable statements) by the use of chi square corrected for continuity. In ten cases the chi square was significant at beyond the .10 level of confidence. In eight of these ten cases the significance of chi square came from an inversion of unfavorable items, while in two cases the significance came from an inversion of favorable items.

Finally, a distribution of these 29 cases into a two-by-two table on the basis of bunching or no bunching (as defined by Wicker's extrapolation of the Thurstone and Chave criterion), on the one hand, and on the basis of a significant number of inversions or lack of a significant number of inversions (using the p values

described above and with a p of less than .10 being considered as significant), on the other, yielded a chi square corrected for continuity of .9 and a phi coefficient of .18. Neither of these values is significant.

DISCUSSION

From the data at hand it is not possible to determine whether the smaller number of irregularities in the five-interval data is a function of the number of intervals or of other factors. In this connection, however, it is interesting to note that a few of the judges complained that 11 intervals were too many.

Since the data reported here are only descriptive in nature, they do not permit a conclusive determination of the causes of these different irregularities. While bunching of items (too many in one interval) and failure to use some intervals have been observed in our data, there is no evidence to explain the causes of these irregularities.

To find an explanation for double markings and omissions, the writer examined all sets of ratings containing these irregularities. This examination showed that omissions were randomly scattered among the items and that in practically every case where double marking occurred, one rating fell on the opposite side of the neutral interval from the other—an observation which suggests that the judges recorded a judgment incorrectly and then failed to erase it before making a second rating. These findings suggest that these irregularities may well be a result of carelessness.

The decision to use inversions as a type of irregularity resulted from the writer's experience with judges who indicated that they were expressing agreement or disagreement with an item rather than making a judgment of it. This type of irregularity seems not to have been discussed by other investigators. The data show tendencies for the same items to be inverted with the same relative frequency by the judges of Thurstone and Chave, of Wicker, and of the writer. These inverted items tend to be predominantly unfavorable statements.

Since 41.5 per cent of all sortings falling in categories 3 and 8 had only one inversion, it is possible that many of these inversions may be attributed to carelessness in recording judgments. But since the judgments of certain items are inverted more frequently and with some degree of consistency by different judging groups than are the ratings of some other items, it does not appear that carelessness provides an adequate explanation for all of them.

Nor does the occurrence of inversions appear to be a function solely of the items. A detailed analysis of the sortings of 10 judges having a significantly large number of inversions showed that in no case were all favorable (or unfavorable) items inverted by any one judge, nor were the judgments of any one item inverted by all judges. In view of these observations it would appear that the inversion process is not a systematic process but may be attributed to an interaction of a particular judge and a particular item.

In view of his experience with judges whose pattern of judgments for the first three items did not follow the expected pattern, the writer proposes as an explanation of inversions the hypothesis that the judge's ego involvement with certain items becomes so strong that it disrupts his set for expressing an unbiased judgment and intensifies a set for expressing an agreement or disagreement. Under the influence of this disruption the judge places his rating at some position other than the one he would place it in if he were judging with an unbiased attitude. If it is assumed that he is judging extreme statements, this displacement would not be noticed as irregular if he displaces his judgment one or two intervals; only when he displaces the statement to such an extent that it falls on the unexpected side of the neutral interval will it be noticed as an inversion as defined here.

While the operation of an ego involvement in this manner may be hard to demonstrate, the possibility of its existence seems consistent with the fact that it is relatively easy to get students to adopt a set for expressing agreement or disagreement with items; but it is difficult to get them to adopt a set for expressing a judgment, since they prefer to express an agreement or disagreement with the items instead.

REFERENCES

1. EDWARDS, A. L., & Kilpatrick, F. P. A technique for the construction of attitude scales. *J. appl. Psychol.*, 1948, **32**, 374–384.
2. HINCKLEY, E. D. The influence of individual

opinion on construction of an attitude scale. *J. soc. Psychol.*, 1932, **3**, 283–296.
3. HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.
4. KELLEY, I. B. The construction and evaluation of a scale to measure attitude toward any institution. *Bull. Purdue Univer.*, 1934, **35**, No. 4, 18–36.
5. KILPATRICK, F. P. A technique for the measurement of attitudes. Unpublished master's thesis, Univer. of Washington, 1947.
6. MILLER, H. E. The construction and evaluation of a scale of attitudes towards occupations. *Bull. Purdue Univer.*, 1934, **35**, No. 4, 68–76.
7. SEASHORE, R. H., & HEVNER, K. A time saving device for the construction of attitude scales. *J. soc. Psychol.*, 1933, **4**, 366–372.
8. SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48**, 135–141.
9. SILANCE, E. B., & REMMERS, H. H. An experimental generalized master scale: a scale to measure attitude toward any subject. *Bull. Purdue Univer.*, 1934, **35**, No. 4, 84–87.
10. THURSTONE, L. L., & CHAVE, E. J. *The measurement of attitude.* Chicago: Univer. of Chicago Press, 1929.
11. WICKER, M. P. A comparison of attitude scale values yielded by scales of differing lengths. Unpublished master's thesis, Univer. of North Carolina, 1950.

Received May 7, 1953.

PERCEPTUAL RIGIDITY AS MEASURED BY ANISEIKONIC LENSES[1]

WESLEY C. BECKER

Stanford University

THE study of personality rigidity has taken a prominent place in the current psychological literature.[2] Cattell and Tiner (6) and Scheier and Ferguson (11) have approached the problem by use of factor analysis. Fisher (8) and others have used a more clinically oriented approach. Though numerous measures of rigidity have been offered, several writers have emphasized the fact that better measures are needed before research in this area can advance. This paper explores the utility of a novel perceptual device, aniseikonic lenses, which seems to avoid pitfalls of other measures of rigidity.

Aniseikonic lenses were initially designed to study distortion in stereoscopic perception. They magnify the image of one eye in the 45-degree meridian and of the other in the 135-degree meridian. The induced effect of the lenses is to make a table appear to tip up like a drawing board or to make a wall lean toward the observer. Preliminary observations by the writer indicated that people varied widely in the time it took to see the effect and in the degree to which distortion[3] was perceived. Some older people could not see any distortion at all. Several six-year-olds reported seeing the effect almost immediately. Individuals in the intermediate age group varied greatly, more or less independently of age. These and other observations suggested that the aniseikonic phenomenon might be related to personality rigidity when the concept is defined, as in Cattell (6), as the ease with which old established habits may be changed in the presence of new demands.

[1] This paper is condensed from a thesis submitted to the Department of Psychology of Stanford University in partial fulfillment of the requirements for the Master's degree. The author wishes to thank Professors Joseph Luft and Clarence L. Winder for advice and assistance throughout the study. He is indebted to Jean Humphrey Block and Wirt Wolff for their assistance in scoring parts of the data.

[2] For general reviews of the literature on rigidity, the reader is referred to Cattell (6) and Fisher (7).

[3] Distortion as used here actually refers to perception more in line with the reality of the situation. But the term is used because phenomenologically the induced effect seems to be a distortion of reality as it is commonly perceived.

General Hypothesis

On the basis of the preliminary observations and as a step towards clarification of the meaning of the observed individual differences, the following hypothesis was made: Individuals who take a longer time and/or report less distortion while looking through the aniseikonic lenses will reveal more rigidity on personality measures than those who take a shorter time to see the effect and/or report a greater amount of distortion.

Measures of Personality Rigidity

Three somewhat different methods of assessing personality rigidity were used. These methods, with a brief description where appropriate, are given below. In addition, the specific predictions that follow from the hypothesis have been indicated.

Ego-control rating. Jack Block (4) and Jean Block (5) have defined a genotypical personality construct called ego control which is rated from a standardized interview. Ego control refers to "the individual's characteristic means of handling or mediating both his internal need-tensions and the demands imposed upon him by the external world" (4, p. 6). In their operational descriptions of three reference points, a number of statements suggest that the overcontroller has difficulty adapting to change and restructuring a situation. The undercontroller easily shifts from one frame of reference to another, while the appropriate-controller's ability to shift with environmental change is intermediate. Some people showing both "over" and "under" tendencies do not fit into the continuum and have been called "bimodal" (5, pp. 29–35).

It was predicted that a shorter time delay and/or greater perceived distortion with the aniseikonic lenses would relate to undercontrolling tendencies, whereas a longer time delay and/or lesser perceived distortion would relate to overcontrolling tendencies.

Rorschach. A number of predictions were made about Rorschach variables assumed to be related to rigidity-flexibility.[4] The predictions have been stated in one direction only to simplify exposition.

It was predicted that a shorter time delay would relate (*a*) to a shorter average initial reaction time, (*b*) to the production of a greater *R*, (*c*) to a greater *Dd*%, (*d*) to the production of more *S* responses, (*e*) to a lower + %, (*f*) to the production of a greater variety of content, (*g*) to the use of a greater variety of determinants, and (*h*) to a higher *Sum C*. It was concurrently predicted that each of the above Rorschach variables would relate to the degree of perceived distortion in a direction opposite to that stated for the time delay.

The Rorschach protocols were scored by Klopfer's method (10) with the exception of form level for which Beck's tables (1, 2) were used. All responses with a form element were included in the + %. Additional white space and color responses were given half weight. The determinant variety score is the number of different determinants used in any way. A different determinant was defined in terms of the groupings under *Total* on the Rorschach Individual Record Blank (9).

Fisher's rigidity scale. Fisher (8) devised a system of weights which are given to various extreme Rorschach scores, the sum of which is assumed to reflect a "general personality rigidity." It was predicted that a shorter time delay and/or a greater estimate of distortion would relate to a lower score on the scale. Since Fisher's weightings were based on Beck's scoring, minor adaptations of the scale were necessary in using it with Klopfer's system. This scale does not add much novelty, because of its overlap with previous predictions.

METHOD

Procedure. Each subject (*S*) was individually given the standardized interview designed by the Blocks (4, 5) from which the rating of ego control was made.

After the rating was completed *S* was instructed as follows:

> Stand here and look at this desk. I'm going to have you put these glasses on in a second, and as soon as you see the desk start to tip, you tell me. I want you to look generally in this direction (center of the table) and try to take in the whole top in one view. Don't move your eyes around freely, but you may have to do some shifting of your glance from front to back to see the tip. Keep this movement to a minimum, however, and try to keep a general fixed focus which takes in the whole top of the desk. Report to me as soon as you can tell which way it has tipped. Say "now" or something like that. I'll keep one eye covered till I start my watch.[5] After the first trial you'll know pretty well what you're looking for.

After *S* had reported the distortion, he was given time for the effect to stabilize and then asked, "how much higher would you say the back is than the front in inches?" The report was recorded. This procedure with the aniseikonic lenses was repeated for seven trials. One minute was allowed to elapse with the lenses off between trials.

Since a preliminary study indicated that learning effects were minimal after the first two trials, an average of the last four out of seven trials was used as the measure of the time delay. The report given on the first trial was used for the estimate of distortion score, since this seemed to represent *S*'s maximum.

In this part of the experiment, *S* stood three feet from the front of a 30-inch-high desk. A two-by-four foot plywood top was used to add texture cues for better depth judgment. On the desk were a flexible neck lamp, a tennis ball, and an oil can about six inches high.

An individual Rorschach was then administered using the procedure set forth by Klopfer and Kelley (10).

Subjects. The *S*s were 10 women and 22 men from an elementary psychology class at Stanford University. Their ages ranged from 18 to 48 with a mean of 24.5 and an *SD* of 7.37. Eleven *S*s who wore glasses underneath the experimental lenses during the experiment had a slightly longer average reaction time, but not significantly so. The total *N* was 32 for all relationships except ego control, where the elimination of 5 bimodals left an *N* of 27.

RESULTS

Reliability and consistency. For reliability purposes, a second ego-control rating was made from the interview protocols by an experienced clinician familiar with the ego-control construct. On a six-point scale, there was agreement within one point in 69 per cent of the cases and a deviation of more than one in 31 per cent of the cases. Discrepancies in ratings were eliminated by a discussion session.

With an analysis of variance technique to estimate the error variance, the reliability coefficient of the averaged time delay score was found to be .95. McNemar[6] has shown the procedure used will lead to values equivalent to those obtained by the Spearman-Brown formula.

[4] The reader concerned with the rationale for these predictions is referred to the original work (3).

[5] The effect of the lenses does not occur with monocular vision. This allows *E* to control precisely the time of starting by covering one eye until both *S* and *E* are ready.

[6] McNemar, Q. Personal communication.

An estimate of the degree of individual consistency of the time delay score in different situations was made in the preliminary study. Four different situations varying in dimension and familiarity of structure were used. By determining the reliability of an individual's mean score based on the four situations, the degree of individual consistency can be estimated. Such an estimate of reliability gives a coefficient of .83. This is based on an N of 15 and is significant beyond the .001 level.

Evidence that the measure is fairly stable over time was obtained on nine *S*s who were retested from three to four weeks later. The rank-difference correlation for the time delay scores was .70. This is a conservative estimate of consistency since the conditions of measurement differed considerably from the first to the second testing.

Relations to ego control. The predictions for both the time delay and the degree of distortion as related to ego control are confirmed. The product-moment correlation between the time delay and ego-control ratings is .58, significant at the .001 level for an N of 27. The product-moment correlation between the estimated distortion and ego control is −.58, significant at the .001 level. The fact that the time delay and distortion scores intercorrelate at −.59 should be kept in mind in interpreting these results.

Rorschach predictions. Tetrachoric correlation coefficients[7] and the chi squares for the 18 predictions to Rorschach variables are presented in Table 1. The tetrachoric coefficients should be interpreted with caution for this small an N. They are presented merely to give an idea of the relative magnitude of the relationships. A shorter time delay with the aniseikonic lenses is significantly related to (*a*) a greater *Dd*%, (*b*) a greater variety of content, (*c*) a greater use of *S*, (*d*) a greater variety of determinants, (*e*) a higher *Sum C*, and (*f*) a lower score on Fisher's rigidity scale. Initial reaction time, *R*, and + % failed to reach the .05 level of significance, although the results were in the expected direction. The degree of distortion was significantly related to *R* and to those same variables which were significantly related to the time delay with the exception of *S*.

[7] All chi squares are based on two-by-two tables. The variables were cut as close to the median as possible. The tetrachoric correlations were computed by means of Thurstone's diagrams (12).

TABLE 1

THE RELATIONS OF TIME DELAY AND DISTORTION TO THE RORSCHACH VARIABLES

RORSCHACH VARIABLE	TIME DELAY			DISTORTION		
	χ^2	p*	r†	χ^2	p*	r†
Reaction time	2.00	.08	.37	1.17	.15	−.29
R	.50	.25	−.18	3.24	.04	.48
Dd%	6.15	.01	−.63	4.40	.02	.53
No. *S*	3.14	.04	−.46	1.89	.10	.36
+%	.50	.25	.18	1.17	.15	−.29
Sum C	8.00	.001	−.71	6.35	.01	.66
Variety of content	6.15	.01	−.63	7.94	.01	.70
Variety of determinants	4.20	.02	−.54	5.87	.01	.60
Fisher's scale	3.14	.04	.46	4.40	.02	−.53

* One-tailed test.
† Tetrachoric coefficient.

SUMMARY AND CONCLUSIONS

On the basis of current theory as to the nature of personality rigidity, it was postulated that persons who revealed such rigidity would show a related perceptual rigidity as manifested by the time delay and degree of distortion perceived when viewing simple structured objects through aniseikonic lenses. Specific predictions were made to Block's ego-control variable and to several Rorschach variables assumed to be related to rigidity.

The data revealed that it was possible to get a reliable measure of the time delay which was quite consistent from situation to situation and over short periods of time. Seven out of the ten predictions for both the time delay and the estimate of distortion were confirmed at the .05 level of significance or better. A shorter time delay was significantly related to (*a*) less ego control, and on the Rorschach to (*b*) a larger *Dd*%, (*c*) the use of more *S*, (*d*) a higher *Sum C*, (*e*) a greater variety of content, (*f*) a greater variety of determinants, and (*g*) a lower score on Fisher's scale of rigidity. The degree of distortion was significantly related to *R* and to the variables just listed with the exception of *S*.

The general hypothesis that the time delay and degree of distortion with aniseikonic lenses

are meaningfully related to other measures of perceptual and personality rigidity is supported by the data.

REFERENCES

1. BECK, S. J. Introduction to the Rorschach method: a manual of personality study. *Amer. Orthopsychiat. Ass. Monogr.*, 1937, No. 1.
2. BECK, S. J. *Rorschach's test.* Vol. I. *Basic processes.* New York: Grune & Stratton, 1944.
3. BECKER, W. C. The use of the aniseikonic glasses as a measure of perceptual rigidity. Unpublished master's thesis, Stanford Univer., 1953.
4. BLOCK, J. An experimental investigation of the construct of ego-control. Unpublished doctor's dissertation, Stanford Univer., 1950.
5. BLOCK, JEAN H. An experimental study of a topological representation of ego-structure. Unpublished doctor's dissertation, Stanford Univer., 1951.
6. CATTELL, R. B., & TINER, L. GHOSE. The varieties of structural rigidity. *J. Pers.*, 1949, **17**, 321–341.
7. FISHER, S. An overview of trends in research dealing with personality rigidity. *J. Pers.*, 1949, **17**, 342–351.
8. FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
9. KLOPFER, B., & DAVIDSON, HELEN. *The Rorschach method of personality diagnosis, individual record blank.* Yonkers, N. Y.: World Book Co., 1942.
10. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique.* New York: World Book Co., 1946.
11. SCHEIER, I. H., & FERGUSON, G. A. Further factorial studies of tests of rigidity. *Canad. J. Psychol.*, 1952, **6**, 18–30.
12. THURSTONE, L. L., CHESIRE, LEONA, & SAFFIR, M. *Computing diagrams for the tetrachoric correlation coefficient.* Chicago: Univer. of Chicago Bookstore, 1933.

Received May 7, 1953.

VALUES AS DETERMINANTS OF WORD-RECOGNITION THRESHOLDS[1]

J. C. GILCHRIST, J. F. LUDEMAN, AND WILLIAM LYSAK
University of Wisconsin

THE issues of motivated perception such as value (8), perceptual defense, and subception (2, 5, 6) are still unclear and the subject of controversy (1, 7). In an attempt to clarify some of the factors involved, the present study was undertaken with the expectation (*a*) that value, both positive and negative, lowers word-recognition thresholds; (*b*) that the effects of value upon such thresholds cannot be explained on the basis of frequency of word usage; (*c*) that negative value should not be assigned equivalent status with emotionality; and (*d*) that the latter raises word-recognition thresholds.

APPARATUS AND PROCEDURE

The major piece of apparatus was an overhead tachistoscopic slide projector. The lamp of this projector was connected with a variable transformer capable of delivering from 0 to 135 volts. A voltmeter giving voltage readings accurate to within ± 1.0% was included in the lamp circuit. The projector was equipped with a shutter which made flash presentations of projected words possible. Flashes were constant in duration (.01 sec.) throughout the experiment. The word image appeared upon a beaded projection screen which had a black ½″ x ½″ fixation mark affixed directly below the site at which the center of the projected word appeared. The screen was located 12 ft. in front of the projector. The words used in the experiment were typed in capital letters on slide mats. Each of the words appeared on two slides; once with the word *ink* typed one space above and below it, and once with *Jew* typed above and below it. These will hereinafter be referred to as "surrounds." The positively valued words were: *able, brave, honest,* and *loyal.* The neutral words were: *chair, frost, orange,* and *salt.* The negative words were: *cheap, dirty, loud,* and *vulgar.* The positive words were selected to be representative of the values deemed positive by Lynd and Lynd (4). The negative words were chosen in the same fashion except that they were also screened to reflect the negative stereotype of the Jew in American society. According to the Lorge-Thorndike word counts (9), the positive words appear in the first, second, second, and fourth thousand most frequently used words, respectively; the neutral words in the first, third, second, and first, respectively; the negative words in the second, third, second, and sixth, respectively. Each set of value words contains 1 four-, 2 five-, and 1 six-letter word.

[1] This study was supported, in part, by funds supplied by The Research Committee, The Graduate School, University of Wisconsin.

The *S*s were 30 University of Wisconsin students enrolled in a first course in psychology. They composed two groups of 15 each which were obtained from the upper and lower quarters, respectively, of a distribution of 235 scores on the Levinson-Sanford (3) scale of anti-Semitism. Possible scores on this scale range from 52 to 364. The group of low-prejudice *S*s had a range from 52 to 98; the high-prejudice group, a range from 199 to 250. None of the *S*s were Jews. The *S*s were not told the reason for their selection, nor their scores on the scale. Despite this, the ratio of high-prejudice to low-prejudice volunteers was much less than one, indicating the operation of some other selective factor in that group. The low-prejudice group was made equivalent to the high-prejudice group with respect to number and date of appearance for the experimental session.

Each *S* was brought into the laboratory and seated in a chair facing the screen which was situated directly in front of the projector. He was given the following instructions:

> This is an experiment in perception. I am going to flash on the screen there (*E* points) groups of three words for just a fraction of a second. The words will be arranged like this. (*E* shows *S* card on which is typed the word *please* with *cat* above and below it.) Your job will be to identify the middle word which will appear just above that little black mark. Before each trial, I'll say "ready" and then one second later the flash will appear. You probably won't be able to see very much for the first few trials but soon the words will become clearer. As soon as you can make them out, tell me what they are. It's all right to guess. Tell me what the word is or what any part of it might be as soon as you can. Are there any questions? O.K. Let's try a few practice words before we begin the experimental words.

Three practice words, *hand, solve,* and *people*—with *bit, and,* and *nod* as their respective surrounds—were used to obtain an initial mean threshold for each *S*. This was done in the following manner. Voltage across the lamp filament was set at 50 volts and the word was exposed once. It was then exposed once at each 2-volt step above 50 volts until the *S* gave two successive correct reports of the surrounded word. Threshold was considered to be the voltage setting at which the first of the two successive correct reports was given. The mean threshold was determined for the three practice words, and 20 volts were subtracted from it. This value was the starting point for each subsequent threshold determination on the experimental words, each of which was given each *S* once in each surround. Each *S*, therefore, appeared in each value and surround cell of the experimental design, with prejudice groups being independent. The threshold determinations with experimental words, except for the starting point, were

TABLE 1
SUMMARY OF ANALYSIS OF VARIANCE

SOURCE OF VARIATION	*df*	MEAN SQUARE	*F*	*p*
Value	2	131.15	9.44	<.001
Prejudice	1	72.83	—*	
Surround	1	28.40	2.04	
Ss within prejudice groups	28	356.49	25.65	<.001
Prejudice × value	2	2.15	—	
Prejudice × surround	1	18.38	1.32	
Value × surround	2	53.57	3.85	<.05
Value × *Ss* within	56	25.02	1.80	
Surround × *Ss* within	28	11.84	—	
Prejudice × value × surround	2	7.64	—	
Residual	56	13.90		

* Tested by *Ss* within prejudice groups.

the same as that described above. The experimental words were presented in a different random order for each *S*. All responses made by the *S* to each exposure were recorded. At the conclusion of the experiment, each S was asked to recall the words seen, and these responses were also recorded. This completed the experimental session.

RESULTS[2]

Since, in the range of values between the lowest starting point of any *S* and the highest single word threshold of any *S*, the relationship between lamp voltage and illuminance was essentially rectilinear, all data are presented in terms of lamp voltage. Inasmuch as no significant difference was found between the preliminary thresholds of the high-prejudice and low-prejudice groups, an analysis of variance was performed on the data, the results of which are presented in Table 1. Individual word thresholds were not used in the analysis. The data used were the *Ss*' mean thresholds for all four words in a value set in each surround. Plots of the data for the significant variables as indicated by the analysis of variance are presented in Fig. 1 and 2. The *t* values for the various points in the two surrounds are all significant at better than the 5% level of confidence: that for the positive mean thresholds yields a $p < .02$; that for the neutral mean thresholds gives a $p < .05$; and that for the negative mean thresholds, a $p < .01$.

Tabulation of prerecognition responses showed no significant differences between groups or between value categories of words when classified in terms of the categories co-valuant, contra-valuant, structural, nonsense, and unrelated used by Postman, Bruner, and McGinnies (8). Nor were there any significant differences found between groups and value categories for recall.

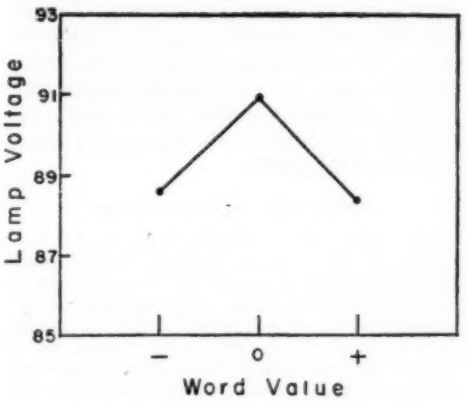

FIG. 1. MEAN THRESHOLDS IN LAMP VOLTAGE OF NEGATIVE, NEUTRAL, AND POSITIVE VALUE WORDS COMBINED FOR BOTH SURROUNDS AND PREJUDICE GROUPS

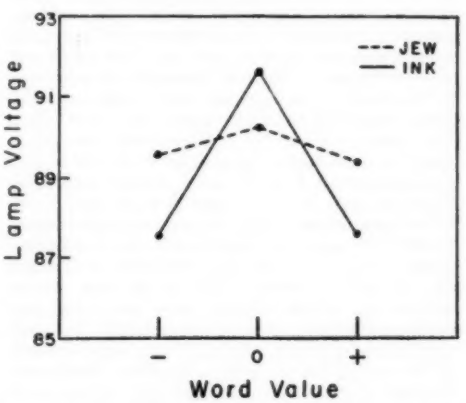

FIG. 2. MEAN THRESHOLDS IN LAMP VOLTAGE OF NEGATIVE, NEUTRAL, AND POSITIVE VALUE WORDS COMBINED FOR PREJUDICE GROUPS

[2] The data reported herein comprise a duplication of a previously unreported experiment, the data from which were discarded when a possible change in projector characteristics was noted at the conclusion of the experiment. The main effects and interactions for which significant differences were obtained are precisely the same in this experiment as in the previous one, although the absolute values differ as a result of the changed apparatus.

DISCUSSION

It is apparent from Fig. 1 that both positively and negatively valued words have lower recognition thresholds than neutrally valued words. That word frequency is an inadequate explanation of this result is shown by the fact that the neutral words have a higher mean frequency than do either the positive or negative words. Further, such a finding indicates that previous investigators have perhaps dealt with only one portion of the curve, or with another aspect of the problem, when they have implied that only positive values lower word-recognition thresholds. McGinnies, in fact, has explicitly stated this latter position: "When . . . individuals are faced with stimulus objects which are not actually threatening, but which represent for them areas of little interest or some antipathy, they also generally display raised thresholds of recognition" (6, p. 244). Our evidence would seem to negate such a statement if by "some antipathy" is meant negative value.

Briefly, we believe that the difference is due to the kinds of negatively valued words used in the several experiments. Words such as *whore* and *bitch* are not just carriers of negative value: saying them in an uncertain social setting (e.g., *S* with an *E* who, at least, is not too well known by *S*) may result in tacit disapproval of the speaker or even overt expressions of condemnation. This position is, of course, that of Howes and Solomon (1) and implies that the phenomenon observed by McGinnies is one of response suppression rather than one of perceptual defense.[3]

It is, however, difficult to maintain this position in the face of the data shown in Fig. 2. Assuming, not unreasonably (10), that our *S*s all saw the surrounds in each condition, it is possible to argue that the effect of the emotionally loaded *Jew* surround actually produced sensitization, as evidenced by the lowered neutral threshold. However, in conjunction with either value category, it produced response suppression as indicated by a reluctance to make the connection of either *brave Jew* or *dirty Jew* until certainty was reached. But why suppression of *brave*—at least by the low-prejudice *S*s?

Nevertheless, it is true that only four of the 30 *S*s ever reported seeing, during the experiment or in postexperimental interrogations, either surround. And in this situation there seems less reason to suspect withholding by *S*s than in McGinnies' case. Even so it is equally, if not more, difficult to maintain the perceptual defense position, for it would appear that from this position one would have to predict raised thresholds for the neutral category in the *Jew* surround. Further, it seems that one would have to predict, as well, an interaction between prejudice, value, and surround, with higher thresholds in the positive category and lower thresholds in the negative category in the *Jew* surround for high-prejudice *S*s, and with the reverse occurring for the low-prejudice *S*s. No suggestion of this trend can be found in the data. In its absence, the results of the experiment remain ambiguous on the question of the effects of emotionality upon word-recognition thresholds. Nevertheless, it is believed that the finding of raised thresholds for the positive category poses problems for proponents of both "response suppression" and "perceptual defense." With the present evidence, we see no immediately clear solution. But in the first flush of reaction against perceptual defense we feel that caution should prevail lest "the baby go the way of the bathwater." There is substantial evidence to the effect that attitudes and value systems selectively change a wide variety of behaviors. This evidence shows that when stimulus or mnemonic material contains values counter to those of *S*, the effect is usually to remove them from *S*'s field. Such evidence indirectly makes some form of a perceptual defense hypothesis plausible. On the other hand, the evidence of this study does seem to indicate that if we want to find out what these effects are in perception, we will have to become more subtle and stop intimidating naive college student *S*s with words like *whore* and *bitch*.

[3] The finding of low thresholds for negatively valued words may also explain Lazarus and McCleary's (2) data wherein they get high GSR's to shock syllables in "pre-recognition" (?) trials but no differences in threshold. The fact that their comparisons were between syllables incorrectly chosen by *S* would make no difference to this argument when their techniques of forced choice and threat of shock are taken into account. At this point, however, we hesitate to make this claim categorically.

SUMMARY

Some effects of word value and emotional context upon word-recognition thresholds have

been investigated using high- and low-prejudice *S*s to define the meaning of the context values.

In general, it was found that (*a*) both positive and negative values lowered word-recognition thresholds in comparison with neutral value, and (*b*) emotionally loaded contexts have the effect of raising the thresholds of both positively and negatively valued words while lowering that of neutral words.

It was pointed out that these data pose problems for both the concepts of "response suppression" and "perceptual defense."

REFERENCES

1. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
2. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
3. LEVINSON, D. J., & SANFORD, R. N. A scale for the measurement of anti-Semitism. *J. Psychol.*, 1944, **17**, 339–370.
4. LYND, R. S., & LYND, HELEN M. *Middletown in transition: a study in cultural conflicts*. New York: Harcourt, Brace, 1937.
5. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: an interim report. *J. Pers.*, 1949, **18**, 171–179.
6. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
7. MCGINNIES, E. Discussion of Howes and Solomon's note on "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 235–240.
8. POSTMAN, L., BRUNER, J., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–154.
9. THORNDIKE, E. L., & LORGE, I. *The teachers' word book of 30,000 words*. New York: Teachers Coll., Columbia Univer., 1944.
10. WHITAKER, E. M., GILCHRIST, J. C., & FISHER, J. W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732–733

Received May 14, 1953.

WORD FREQUENCY AND THE MEASUREMENT OF VALUE AREAS[1]

DONALD R. BROWN AND JOE ADAMS

Bryn Mawr College

THE measurement of value areas by such tests as the Allport-Vernon Scale of Values (2) and the recent revision by Lindzey (3) has been characterized by Solomon and Howes (7) as a function of word frequency. They reach this conclusion as a result of their critical analysis of the earlier work of Postman, Bruner, and McGinnies (6) on the relationship between value and perceptual threshold. We have already treated the Solomon and Howes frequency hypothesis in a theoretical manner (1). The present report is the result of an attempt to test our contention that such constructs as "value" or "meaning" are useful theoretical constructs, admissible to operational demonstration, and necessary to explain results obtained on such tests as the Allport-Vernon. We decided to test the Solomon-Howes word frequency hypothesis by maximizing conditions in favor of this hypothesis, so that a clear-cut prediction on the basis of the word frequency hypothesis could be made. Our prediction, on the other hand, was that central cognitive mechanisms or structures would tend to overcome wide differences in the frequencies of the stimuli.

METHOD

Six forms of a value category test were constructed after the general pattern of the Allport-Vernon (2) test, using the six Spranger value categories. The test consisted of 40 multiple-choice items. Each item contained three one-word choices, and each choice represented one of the six value categories. Since there are 20 combinations of 6 forms taken 3 at a time, the 40 items included each possible triple combination twice. Each of the 6 forms favored one of the value categories in terms of frequency; that is, each time that category appeared in the form as a choice, it was represented by a very high frequency word from the Thorndike-Lorge count (9), while all other value categories on that form were represented by very low frequency words. The high frequency words (of which there were 120 in all, 20 for each value category) were chosen from the G or general count of Thorndike and Lorge and were words which in 57 cases appeared 100 times or more in a million words; in 47 cases appeared from 50 to 99 times in a million words; and in 16 cases appeared from 20 to 43 times. The infrequent words appeared 10 times or less in a million words, with the great majority of them having fewer than 5 appearances. Occasionally, it was necessary to go into lists other than the G count to get the right word, but always the most infrequent word that could be found that connoted the proper value category was chosen.

The following item is typical of the test, only one of the two words (high and low frequency) for each of the three value areas actually appearing on a given form:

1. Scientific research should seek

(Theoretical)	Truth. (AA)[2]	Verity. (2)
(Social)	Social gain. (A-AA)	Humanitarianism. (2)
(Economic)	Profit. (A)	Remuneration. (1)

The subjects (*Ss*) were 121 women students in the required history of philosophy course of a liberal arts college. They were divided into six groups of approximately 20 each and were given the Allport-Vernon-Lindzey Scale of Values (3),[3] followed immediately by one of the six forms of our test.

RESULTS

Table 1 shows the mean score and the standard deviation (in parentheses) for each of the six groups of *Ss* for the six value categories. The number of *Ss* in each group is indicated in parentheses beside the letter corresponding to the form which that group took. Each form is named by the letter of the category favored (in word frequency) by that form.

According to the word frequency hypothesis, the diagonal means should be larger than other means in the same column. Table 2 shows the mean of the group taking the favored form, the mean of the other five groups, and a t test of the difference between the two means for each value category. The p value is that for a one-tailed instead of a two-tailed test, as the direction of the difference is predicted by the

[1] Part of this paper was delivered at the Eastern Psychological Association meetings in April 1953, at Boston, Massachusetts.

[2] The following is a key to the symbols appearing in parentheses after each word. These symbols did not, of course, appear on the test.

AA—100 or over occurrences per million words.

A—50 or over occurrences per million words.

The numerals refer to the actual number of occurrences per million words.

[3] Since the publication of the Solomon-Howes article, the Lindzey revision of the Allport-Vernon was published. The changes between the old and new forms are not relevant for our purposes and, therefore, we used the Allport-Vernon-Lindzey revision in our experiment because it was the only form of this test available.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF SCORES OF SIX DIFFERENT GROUPS ON THE SIX VALUE CATEGORIES

FORM	VALUE CATEGORY					
	AESTHETIC	ECONOMIC	POLITICAL	RELIGIOUS	SOCIAL	THEORETICAL
A(21)	*10.05*	4.00	6.10	5.67	8.24	5.43
	(3.7)	(2.2)	(3.1)	(4.1)	(3.9)	(2.5)
E(22)	8.73	*3.18*	7.18	5.00	8.18	6.91
	(2.8)	(2.2)	(3.2)	(4.0)	(2.8)	(3.2)
P(20)	9.05	3.15	*5.65*	6.65	7.90	7.20
	(3.0)	(2.2)	(2.2)	(5.3)	(3.2)	(2.1)
R(19)	7.42	3.37	7.00	*8.05*	7.95	4.95
	(3.8)	(2.2)	(3.7)	(4.2)	(2.7)	(3.3)
S(20)	8.80	3.25	6.15	5.55	*9.25*	6.45
	(3.0)	(2.5)	(2.1)	(4.1)	(3.2)	(2.7)
T(21)	7.95	3.05	6.24	3.62	9.24	*9.67*
	(3.7)	(2.4)	(2.0)	(3.5)	(3.4)	(3.1)

TABLE 2

DIFFERENCES BETWEEN MEANS ON FAVORED FORM AND OTHER FIVE (NONFAVORED) FORMS FOR EACH VALUE CATEGORY

VALUE CATEGORY	MEAN, FAVORED	MEAN, NONFAVORED	DIFF.	t	p (1-tailed)
Aesthetic	10.05	8.40	1.65	1.83	<.07
Economic	3.18	3.37	−0.19	−0.35	
Political	5.65	6.53	−0.88	−1.49	
Religious	8.05	5.28	2.77	2.59	<.01
Social	9.25	8.31	.94	1.18	
Theoretical	9.67	6.21	3.46	4.68	<.01

word frequency hypothesis. The value of t is in each case computed so as to eliminate differences between the means of nonfavored forms as a contribution to the estimate of sampling error; this means that we are not considering the five nonfavored forms as a random sample from a universe of nonfavored forms. If this correction had not been made, each t would be *smaller* than it is as actually computed.

By computing t in the way described and by making a one-tailed test we have given the word frequency hypothesis every chance of being supported. As can be seen from the table, however, scores on only two categories, religious and theoretical, are significantly higher. On two of the categories, economic and political, the obtained differences are not even in the predicted direction. Our results, therefore, certainly do not support the hypothesis that scores on the Allport-Vernon test or even a modified version of it are a function solely of word frequency. We would argue further that our results do not support even the less extreme contention that word frequency per se is a variable that enters into the determination of the value score. We do not, of course, deny that the frequency with which an individual uses or perceives a word is correlated with his degree of interest in the area related to the word; in fact, interest may well be the determiner of frequency.

On the other hand, since two of the six differences are clearly statistically significant and a third one approaches significance, we feel obligated to account for these differences in some way that will not imply that all six differences should have been significant. Since our division of the 121 *S*s into six groups was not entirely random, some sections systematically getting more of a given form than others, and since we had Allport-Vernon-Lindzey test scores for all our *S*s, we made an analysis of covariance for each of the six value areas, eliminating the differences between forms attributable to regression of scores on our test on the Allport-Vernon-Lindzey scores. Following the procedure given by McNemar (5) and Edwards (4), we obtained the results presented in Table 3. As can be seen from this table, differences among forms on aesthetic, religious, and theoretical are statistically significant at the .05 level or better.

We can believe that the reason for these differences is the fact that a change in word frequency often produces a change in meaning. The word "poesy," for example, in addition to being less frequent than the word "literature" has an obviously different meaning. In some

TABLE 3

RESULTS OF THE ANALYSES OF VARIANCE AND COVARIANCE FOR THE SIX CATEGORIES

CATEGORY	F (ANALYSIS OF COVARIANCE)	p	F (ANALYSIS OF VARIANCE)	p
Aesthetic	3.44	<.01	1.40	
Economic	.80		.45	
Political	.94		.85	
Religious	2.73	<.05	2.39	<.05
Social	1.51		.71	
Theoretical	6.59	<.001	6.60	<.001

Note.—Each F is the ratio of the variance estimate based on variation of form means to the variance estimate based on variation within groups. The degrees of freedom are 5 (numerator) and 117 (denominator).

TABLE 4

CORRELATIONS BETWEEN SCORES ON THE ALLPORT-VERNON-LINDZEY TEST AND SCORES ON THE SIX DIFFERENT FORMS OF THE MODIFIED TEST

FORM	CATEGORY (ALLPORT-VERNON-LINDZEY)					
	A	E	P	R	S	T
A	.72	.66	.53	.88	.63	.47
E	.46	.76	.62	.71	.72	.42
P	.70	.58	.48	.72	.61	.49
R	.53	.71	.38	.88	.43	.75
S	.68	.69	.13	.68	.56	.55
T	.70	.56	.29	.79	.54	.74

cases, like the above, less frequent words have an archaic or unattractive connotation or are unknown. Although we made an effort to keep meaning relatively constant, we found it impossible to do so.

Further, it seems to us that with a more careful choice of low and high frequency words, with special attention to their connotations, than we made in the construction of the various forms of our test, scores on the category favored in terms of frequency could be made consistently lower than scores on that same category on other, nonfrequency favored forms.

Our central cognitive interpretation of value and interest scores is further favored by a correlational analysis of our data. Table 4 presents correlations between scores on the six forms of our test and the Allport-Vernon-Lindzey for all six categories. Within a given column the correlations are completely independent, being based on six different groups. Within a given row the correlations are based on the same group and therefore are not completely independent. The fact that the coefficients are consistently positive and, with a few exceptions, not significantly different in size, is in accord with a central cognitive interpretation of value scores and not at all easy to explain or predict in terms of a word frequency hypothesis.

DISCUSSION

It appears from the above results that while frequency is undoubtedly important in determining perceptual thresholds for verbal material, it alone does not determine performance on value tests. One might question the crucial nature of our results by raising the criticism that individuals, on the basis of their experience, do differ in the amount of commerce that they have with certain words. Our answer to this criticism is twofold: (*a*) A relationship was shown by Howes and Solomon (7) between duration threshold and Thorndike-Lorge count when a group of *S*s was used; if the Thorndike-Lorge count is invalid when as many *S*s are used as in our experiment, their results are unexplained. (*b*) The differences between the high frequency words and low frequency words used in our experiment were so great that it seems implausible that, with the number of *S*s used, differences in frequency of individual usage would not be appreciable.

However, we do not deny the importance of individual differences in word frequencies. Word frequency is not the causative agent in the results of perceptual threshold experiments dealing with verbal material, but rather a reflection of the past experience of the *S*, and therefore will be correlated with measures such as the Allport-Vernon-Lindzey test or our own forms of that test. Individual word frequency then becomes merely another indicator of some central mechanism, such as values, attitudes, interest, or "style of life," which is the causative factor for both probability of word usage by that individual and his performance in situations designed to reflect his basic belief system.[4]

SUMMARY

Six forms of a value-area test were constructed after the general pattern of the Allport-Vernon to test the Solomon-Howes hypothesis that Allport-Vernon scores could be explained as a function of the log frequency of the answer choices. Each form favored frequency-wise one of the six Spranger value areas. One hundred twenty-one women enrolled in a history of philosophy course were tested in six groups; each group received one form of our test and the Allport-Vernon-Lindzey test.

1. The test of the Solomon-Howes frequency hypothesis failed to show consistent changes in scores on the given value area favored for each group, compared to the five other groups in that value area in our test, as would be predicted by the frequency hypothesis.

2. Correlations of our test with the Allport-Vernon-Lindzey scale remain significantly posi-

[4] We have reason to believe from private communication that Solomon and Howes may be in basic agreement with us on this point.

tive even with frequency changes, as would be predicted by the postulation of a central cognitive affective construct which may be called "value area."

REFERENCES

1. ADAMS, J., & BROWN, D. Values, word frequencies, and perception. *Psychol. Rev.*, 1953, **60**, 50–54.
2. ALLPORT, G. W., & VERNON, P. E. *A study of values.* Boston: Houghton Mifflin, 1931.
3. ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *A study of values.* (Rev. Ed.) Boston: Houghton Mifflin, 1951.
4. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.
6. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–155.
7. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.
8. SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952. **43**, 195–201.
9. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* (1st Ed.) New York: Teachers Coll., Columbia Univer., 1944.

Received May 15, 1953.

THE ASSESSMENT OF GROUP OPINION BY LEADERS, AND THEIR INFLUENCE ON ITS FORMATION

GEORGE A. TALLAND

Institute of Psychiatry, Maudsley Hospital, London

IN A recent study Chowdhry and Newcomb (1) tested the relative abilities of leaders in estimating group opinion and found strong evidence to confirm their hypothesis that "leaders should be more accurate than non-leaders in their estimates of the attitudes of other members toward issues relevant to the group's interest." The authors advanced two complementary explanations to account for their findings. Leaders may be accorded their status because of their superior ability in assessing group opinion, and their more accurate evaluation may be due to their position which furnishes them with better opportunities for assessing the attitudes of their followers. The latter was not considered as sufficient explanation in itself.

While Chowdhry and Newcomb did not indicate that only these two circumstances could account for their results, they certainly omitted considering a third which, in their experimental setting, appears equally probable. Leaders may know best the opinions of their groups because they, more than any other member, were influential in formulating these opinions. This line of reasoning would also apply to leaders more generally. Everyday experience in industry, in the services, in political and social organizations, as well as history, demonstrates the fact that leaders are not invariably in close touch with the opinions of their followers. Action leaders in a battle or mob situation, for instance, are likely to assume that role exactly through their insensitivity to the attitudes of those they impel with their commands. The purpose of the present report is to support the argument that leaders know best the opinions of their groups because by their prestige, authority, or other influence they bring their followers in line with their own views.

METHOD

The experiment formed part of a series of psychometric studies of therapeutic groups at the Maudsley Hospital, London.[1] Data collected for the purposes of other researches were utilized, and only a minor modification of the testing procedure was introduced in order to create experimental conditions for the present research.

The applicability of Chowdhry and Newcomb's findings to this type of group was first tested on ten psychotherapeutic groups, totaling 64 members. These had been in treatment for varying periods of from 2 to 18 months, and each consisted of five to eight members, some being all male, others all female, and others of both sexes.

As in the sociometric procedure, leadership status was based on one's fellow group members' assessments. The information, however, was not that of choices and rejections on specified criteria but complete rankings of all other group members. Group scales were derived from the summation of individual ranking scales. This procedure was designed by F. K. Taylor (7), two of whose status scales were employed in the present study, *public popularity* and *dominance*. The individual rankings from which the popularity scale is derived are given in answer to a question: "... estimate the general popularity of each member in the group ... you should not allow your own feelings towards a particular member to influence your judgment." The criterion of dominance rankings was the degree of a member's domination of, and influence upon, group discussion, irrespective of the helpful or harmful effect of this influence. To these was added a new scale, *leadership role*, derived from rankings of the members according to the leadership they had displayed in the group.

So as to provide for a common metric irrespective of the size of the group, all group summated scales were converted according to Taylor's (7) formula into t_p scores, these being $100 \frac{t - (n - 1)}{(n - 1)(n - 2)}$, in which t is the sum of all rank scores assigned to a group member by his colleagues, and n, the number of members in the group. The range of these scores is from 0 to 100, with a mean of 50. Any two or more t_p scales can be summated to provide composite scales.

In order to allow for the highest possible degree of comparability with Chowdhry and Newcomb's findings, leadership was measured by a composite scale which combined direct rankings for leadership role, dominance, and popularity. These are very similar to Chowdhry and Newcomb's first, second, and fourth sociometric criteria. No scale equivalent to their third criterion (worthiness of representing the group in a convention) could be obtained from this type of group. Apart from theoretical considerations which suggest that in this type of group, too, functional criteria of leadership are preferable to sociometric criteria (6), the composite scale also gave rise to some practical disadvantages. In one group no leader emerged, as the lowest score on the composite t_p scale was above 90, and thus too close to the mean to justify the assignment of leadership status to its holder. When judged by the composite scale, this was a leaderless group, although dominance is one's criterion, if the group had a leader

[1] Carried out with a grant from the Bethlem Royal Hospital Research Fund and a University of London Postgraduate Studentship.

scoring as low as 12.5 who, however, had high scores on the popularity and leadership role scales. (The data being rank scores, low scores indicate high status.) In another group two members tied for leadership on the composite scale, and two others scored so high (above 255) that both ought to be considered as bottom scorers, or "isolates."

Throughout the report, leadership will be defined as top score on the composite scale if this is below $t_p = 90$ which would be equal to $t_p = 30$ on a single scale, and isolation as bottom score if higher than $t_p = 210$. As an alternative measure, top and bottom scores on the dominance scale will also be considered as indices of functional leadership and its opposite.

Group opinion was probed in one of the few areas which could be safely assumed to be of common interest to all members within a group: the therapeutic usefulness of various topics of discussion to the group. A list of 15 topics was submitted to the subjects (*Ss*), each of whom was asked to rank these individually with a view to their usefulness to the group as a whole. Concordances of rankings within each group were significantly higher than chance at the 1 per cent level of confidence as measured by Taylor's (7) modification of Kendall's (4) *W* coefficient. In view of this, a measure of group opinion could be obtained by summating the individual rank scales. This summated scale was ranked, and accuracy of evaluating group opinion was measured by the rank correlation coefficient between each member's individual rank scale and the ranked summated scale. The coefficient used was rho. The applicability of Chowdhry and Newcomb's conclusions to the therapeutic groups was tested by examining the positions of the most and least accurate assessors of group opinion on the composite scale of leadership and the dominance scale of each group.

Up to this point ten groups provided data for the research. The final experiment, however, could be applied to seven groups only, with a total membership of 44. The topics were ranked by each member on the basis of (*a*) personal helpfulness, and (*b*) helpfulness to the group. These were submitted, and the groups were then asked to furnish a collective rank scale which would express the opinion of the group as a whole. In order to provide this information, the groups were told to discuss the matter, and the experimenter (*E*) left them to proceed with the task. Three of the ten groups were not available for this experiment, and in one of the remaining five groups two members could not attend the discussion. Neither of these occupied extreme positions on either the composite leadership or the dominance scale.

RESULTS

Nine leaders and nine "isolates" emerged from the ten composite scales. In one group the lowest score was above 90, in another the highest was below 210. None of these nine leaders had the highest assessment coefficient in his group, nor did these leaders tend to score in the clusters of best assessors, which are an artifact of the method by which group opinion was obtained. Two leaders made the least accurate evaluation of group opinion in their groups, and one of these leaders tied for top score on the composite scale. Two isolates gave the least accurate, but two others gave the most accurate, evaluation in their group. None of the most dominant members surpassed all his group colleagues in accuracy of assessment. Two gave the least accurate evaluation in their group, as did two of the least dominant members. Two bottom scorers on dominance made the best assessment of group opinion. It should be noted that each group had at least one member who scored below $t_p = 30$ and one who scored above $t_p = 70$ on the dominance scale. Assessment coefficients in the ten groups are set out in Table 1, giving details of evaluating the summated group opinions. Leaders and isolates on the composite scale are marked L and I, the most and least dominant members D and S. The bar over L means a tie.

Replacing the summated scales of group opinion by the collective rankings we find that the association between leadership and accuracy of assessment becomes clearly marked. Five of the six leaders surpass their fellow group members in evaluating the collective opinion. Only two of the five isolates gave the least

TABLE 1

COEFFICIENTS OF ASSESSMENT OF GROUP OPINION BEFORE PUBLIC DISCUSSION

I	II	III	IV	V	VI	VII	VIII	IX	X
.89	IS .88	.82	.77	.89	.84	.93	IS .83	.80	.90
.88	.80	.77	S .71	IS .88	.83	.88	.82	.76	.63
.82	.68	IS .73	.48	LD .85	LD .76	.85	LD .72	.71	.55
.81	.58	.58	D .43	.80	.72	.80	.64	LD .69	$\bar{L}$ .41
LD .71	LD .57	LD .57	I .42	.78	.68	.78	.58	.65	IS .33
.68	.29	.54	.40		IS .40	.69	.35	S .63	$\bar{L}\bar{D}$ .29
.64		.19				IS .63			
IS .62						LD .48			

TABLE 2
COEFFICIENTS OF ASSESSMENT OF GROUP OPINION AFTER PUBLIC DISCUSSION

I	II	III	IV	V	VI	VII
LD .75	LD .75	LD .69	D .55	LD .93	LD .86	.85
.69	.62	IS .64	.41	IS .80	.84	.78
.59	.57	.63	S .32	.69	.78	.77
.43	IS .45	.32	.31	.69	.53	.74
IS .23	.30	.17	.27	.61	.50	.64
−.40	.05	.12	I .16		IS .24	IS .55
		−.02				LD .44
						.36

accurate assessment. Six of the seven top scorers on dominance gave the best assessment of collective opinion, only one bottom scorer the worst. The new assessment coefficients are set out in Table 2.

DISCUSSION

The results confirm the hypothesis that leaders influence the formation of group opinion when this emerges in the course of discussion, and that this influence is likely to account to some extent for the finding that leaders make the most accurate evaluations of group opinion. Before the matter at issue was publicly discussed, not one of the nine leaders in the ten groups had assessed the group opinion more accurately than had other members. One group was leaderless according to the criterion adopted.

After the question had been publicly discussed, i.e., the psychotherapeutic groups had become discussion groups in the commonly accepted sense of the term, all but one of the six leaders or seven top scorers on dominance were found to have assessed the group's collective opinion more accurately than any of their colleagues. The one exception to this rule may be accounted for as a chance occurrence, though it can be explained by the unusually high agreement among the individual opinions of the four members who followed the leader on the composite scale. With correlation coefficients ranging from .80 to .93 they established a group norm which hardly changed in the course of discussion, and from which the leader had markedly deviated. The results throw little light on the isolate, except that in this situation he does not tend to assess group opinion less accurately than other members. The change in the accuracy of the leader's assessment of group opinion is highly significant. It would be difficult to account for it otherwise than by his direct influence on the shaping of group opinion, whether this influence is exerted by deliberate pressure or by prestige, because of more forceful advocacy or more clearly formulated views on the question at issue.

The departing point of the present experiment was Chowdhry and Newcomb's research. The groups examined satisfied three of their four criteria. They were organized around a definite interest pattern, they provided a basis for face-to-face interaction and for phenomena of leadership to emerge, and their members were sufficiently familiar with each other to be able to evaluate group opinion. Leadership, too, was defined on criteria similar to those adopted by Chowdhry and Newcomb. Admittedly, the presence of the therapist did not allow for a formal recognition of leadership, but this role did emerge in a functional sense. The principles of the therapeutic technique employed (2, 8) prescribed the minimum of interference by the therapist, and, in fact, it was found (5) that the latter's personal judgment had little influence on the shaping of his patients' opinions as far as the usefulness of the topics of discussion is concerned. While psychotherapeutic groups are not meant to be representative of groups in general, nor rankings of discussion topics for their usefulness a typical example of collective opinion, there is no reason to doubt that what has been found in these groups would also occur in others. If the influence of leaders in shaping group opinion is so marked when their status, far from being officially recognized, is established under the shadow of an appointed authoritative conductor, it is likely that official leaders who discharge their functions as a matter of duty would exert at least as great an influence.

Also, if this influence is so effective, after each member had first to form his opinion independently on the same issue and has therefore some motivation to resist prestige suggestion, it may be expected that the leader's influence would be at least as effective when opinions in a group are initially formed by means of public discussion.

The conclusions drawn from this study are intended to complement those of Chowdhry and Newcomb and the discordant findings of Hites and Campbell (3). The results suggest that the leader's more accurate evaluation of group opinion need not be attributed primarily to his more sensitive perception of attitudes and atmosphere in the group, nor to this factor combined with his greater opportunities of communication. They also suggest that publicity rather than familiarity is the operative attribute of opinions in this context. In order to discover the circumstances by which leaders arrive at more accurate evaluations of group opinion than do followers, investigations carried out after such opinions had been formulated by public discussion fail to assess the leader's influence on the outcome of these discussions. Concerning Chowdhry and Newcomb's second hypothesis, i.e., that the leaders' more accurate assessment of group opinion may be due to their better opportunities for communication, though this factor would not in itself account for the finding, the present experiment gives support to the qualifying clause. Theoretically all members of a therapeutic group have the same opportunities of communication. In fact, leaders communicate more frequently and more explicitly to others and receive more communications from their fellow group members. It would therefore have been expected that their views should be most influential on judging group opinion before its explicit formulation. The results did not confirm this expectation.

SUMMARY

Leadership status was defined and established for each member of ten psychotherapeutic groups on a combined criterion of *dominance*, *popularity*, and *leadership role*.

By asking the members of these groups to rank 15 topics of discussion in the order of their usefulness in therapy, group opinions and individual evaluations of these group opinions were obtained on a matter of common interest.

It was found that leaders were not more accurate in assessing group opinion than were other members, before a public discussion in the course of which group opinion was formulated.

When groups were asked to discuss the relative usefulness of these topics publicly and to present a collective ranking, it was found that the collective opinion was, in six cases out of seven, closer to the leader's individual opinion than to those of other members.

The results confirm the hypothesis that leaders influence the formation of group opinion, bringing this in line with their personal views. It is suggested that this influence should be considered in accounting for the finding that leaders evaluate group opinion more accurately than other members.

REFERENCES

1. CHOWDHRY, K., & NEWCOMB, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 51–57.
2. FOULKES, S. H. *Introduction to group-analytic psychotherapy*. London: Wm. Heinemann Medical Books, 1948.
3. HITES, R. W., & CAMPBELL, D. T. A test of the ability of fraternity leaders to estimate group opinion. *J. soc. Psychol.*, 1950, **32**, 95–100.
4. KENDALL, M. G. *Rank correlation methods*. London: Charles Griffin, 1948.
5. TALLAND, G. A. Leadership in psychotherapy groups. Paper read at the Conference of the Brit. Psychol. Soc., 1953.
6. TALLAND, G. A. Étude du leadership dans les groupes de psychothérapie. *Bull. de Psychol.*, **7**, 224–228 (Paris: Sorbonne), 1954.
7. TAYLOR, F. K. Quantitative evaluation of psychosocial phenomena in small groups. *J. ment. Sci.*, 1951, **97**, 690–717.
8. TAYLOR, F. K. On some principles of group therapy. *Brit. J. med. Psychol.*, 1952, **25**, 128–134.

Received May 18, 1953.

PREDICTION IN THE CLINICAL METHOD AND INTERRELATIONS OF BIOCHEMISTRY, PSYCHIATRY, AND PSYCHOLOGY[1]

ELI ROBINS AND IVAN N. MENSH

Washington University School of Medicine

AT THE 1950 annual meeting of the Canadian Psychological Association, C. R. Myers (29) took clinical psychology to task for too little use of the "prediction-test" method in clinical work. The scientific method of hypothesis-prediction-verification falls by the wayside when only the first step is taken, and Myers was concerned that "in the clinical setting, we are forever taking the first step—guessing—without being able to (or perhaps without daring to) follow it up with the next two steps—predicting and verifying." His search of the *Psychological Abstracts* for the years 1947–1949 revealed over 16,000 titles, but only 24 were listed under the entry "Prediction." Of the latter, "16 appeared in 1947, 5 in 1948, and only 3 in 1949. It would not appear that they are increasing in frequency either." Most of the 24 studies referred to the prediction of success at school, college, in marriage, or on parole; two used the clinical interview method for predicting success in pilot training; and only two were "more strictly 'clinical' studies."

Among the reasons Myers sees for the infrequent use of prediction are (*a*) "the lurking suspicion that it is really not possible"; (*b*) group- rather than individual-centered statistics (26, 45); and (*c*) "the ordinary practical difficulties of using prediction effectively as a routine procedure." Further, if prediction is indeed the criterion of validity in science, "and if validation of current theory and practice is the most urgent need in clinical psychology, it is high time we made more deliberate use of prediction as a means of increasing the precision of our clinical understanding." Suggestions for clinical predictions were made by Myers and included the routine use, not only in research, of "the spiral pattern of observe-guess-predict-check-observe-guess-predict-check"; explicit statements, in definite and quantitative terms, and records of clinical predictions in a form which may later be checked by others; actual checking of predictions against outcome and examination of errors in prediction; and short-term predictions, which are easier, more profitable, and allow more prompt checking and revision.

Psychologists have not been entirely remiss, however, in studies of prediction. In May 1946 Malamud (25) summarized a number of studies designed to predict response to shock treatment; last year Rosenzweig (36) suggested a method of validation by successive clinical predictions; the Minnesota group developed an empirical scale of the MMPI (10) for selecting patients who respond successfully to convulsive shock therapy; statistically-oriented clinicians and clinically-oriented statisticians have explored new methods for treating clinical data (26, 45); and for some years Wittman (44) has, with her Prognostic Rating Scale, studied prognosis in hospitalized psychotics. Also, as Myers noted in his presidential address (29), Horst's monograph "The Prediction of Personal Adjustment" (15), and Sarbin's review of "The Logic of Prediction in Psychology" (37) present relevant lessons and rationale.

The study of validation and prediction is not alone a concern of the psychologist, and other students of human behavior have turned to research in this area. Biochemistry, psychiatry, and other branches of medicine now are investigating predictive indices, as witness the Cornell group's "A Clinical Guide to Prognosis in Stress Diseases" (3), Kline and Tenney's study of Sheldon's somatotypes (19), and stress studies (39) in various diagnostic groups. For example, the ultimate aim of biochemical studies concerned with the prediction of behavior is to be able, from a single chemical finding or a pattern of chemical findings, to predict important and relevant behavioral characteristic(s) of the organism in question. The present paper gives a number of examples of the prediction of behavior in the clinical situation utilizing the interrelation of biochemical and physiological techniques, and their psychological and psychiatric correlates.

[1] This paper is one of a series of papers in a symposium presented at the 1951 annual meeting of the American Psychological Association.

These are selected to illustrate the contribution of predictive methods from the fields of biochemistry, psychiatry, and psychology. Before discussing any detailed experimental work, two examples will be presented to point out some of the problems involved in prediction. First, behavioral changes occurring in a severe hypoglycemic reaction in relation to a low blood sugar, resulting in coma, can be predicted with almost complete certainty. The individual in question will have seriously disturbed brain function manifesting itself in disorientation, confusion, and marked disturbances of behavior leading to coma. This is a very precise prediction in terms of the relation of blood sugar level to behavioral disturbances. It is also an extremely important one, both in terms of therapy and in the inferences which may be drawn about metabolism in the brain. In this example one makes a nearly exact *individual* prediction.

The second example concerns the relation of the level of blood lipids to probability of recovery from a schizophrenic or manic-depressive psychosis. As will be discussed later, it is known that, in general, high blood lipid level is associated with a high probability and low blood lipids with a low probability of recovery from either of these diseases (13). However, while the relationship between high blood lipids and recovery is highly significant, in schizophrenia only 55 per cent of those with high blood lipids were well five years after they were first examined. Consequently, for any given schizophrenic patient with high lipid level one cannot predict whether or not he will recover, although the probability is much greater than if he had a low lipid level. This is in sharp contrast to the first example of extremely low blood sugar in which practically every case affected could be predicted. Since a 100 per cent predictive accuracy is rare, the problem that arises is one of evaluating the usefulness and significance of the tests that can give an answer only in terms of more probable or less probable.

There is a second major problem: What kind of behavior are we going to predict? In the first example of low blood sugar, an almost perfect predictive accuracy was achieved if we defined the behavior to be predicted as disorientation, confusion, and marked disturbances in behavior leading to coma. On the other hand, if we wished to predict whether the patient during his behavioral disturbance will hallucinate, dance a jig, curse the boss, lose his temper, or speak lovingly to someone he ordinarily hates (all these behaviors have been observed in hypoglycemic reactions), we cannot do so accurately from the level of the blood sugar. We must, therefore, be careful to define exactly what we wish to predict.

Various experiments concerned with different kinds of prediction will be discussed. These kinds of prediction include: (*a*) Does a given test predict that a certain disease or behavior not present now will occur in the future? (*b*) Does a given test predict that a certain disease or behavior is present at the time of the test? (*c*) Does a given test predict recovery or failure to recover from a given condition? (*d*) Can we predict the effects of lack of or excess of a given drug or other chemical substance? It may be seen that diagnosis and treatment are involved here as well as prognosis.

TOXIC AND NUTRITIONAL FACTORS IN BEHAVIOR

Table 1 summarizes the behavioral effects of a lack of specific nutritional factors and of an excess of certain drugs. These effects are variable, but they generally lead to psychotic behavior. What does this mean in terms of prediction? It can be predicted that the majority of individuals on an experimentally

TABLE 1

THE EFFECTS OF SOME TOXIC DRUGS AND LACK OF CERTAIN VITAMINS ON BEHAVIOR IN MAN

SPECIFIC VITAMIN LACKED	BEHAVIORAL RESULT
Thiamin	Fatigue, nervousness, work inefficiency, palpitation
Niacin	Delirious psychosis*
Vitamin B_{12}	Delirious psychosis
SPECIFIC DRUG ADMINISTERED	**BEHAVIORAL RESULT**
Bromides	Delirious psychosis
Atropine	Delirious psychosis
Mescaline	Hallucinations, delusions, anxiety, finally delirious psychosis
Alcohol	Inebriation

* Delirious psychosis = disorientation, confusion, delusions, hallucinations, unusual behavior, and (in some cases) stupor.

deficient thiamin diet will develop the indicated symptoms (17, 42, 43), that a very small percentage will develop behavioral symptoms of B_{12} lack (1) (assuming the brain lesions of pernicious anemia are due to vitamin B_{12} lack), and that an appreciable but unknown percentage will develop symptoms of nicotinic acid lack (6, 18). Further, early replacement of the missing substance will relieve the symptoms in most cases. In addition, these findings are important for research in that they give a lead concerning the chemical metabolism of the brain. For example, all the vitamins mentioned are important coenzymes in carbohydrate metabolism (23). This is, therefore, another bit of evidence concerning the crucial importance of carbohydrate metabolism in brain function.

In the second group of cases, the majority of subjects will develop symptoms after a sufficient dose of bromides (21), alcohol (14), atropine (20), or mescaline (40). It is also true that a certain small number of subjects will develop symptoms on a low dosage of these drugs. However, since little is known of the mechanism of action of these drugs, we cannot predict which subjects will be the ones to develop symptoms on a low dosage, but can only predict that the great majority of subjects will develop symptoms if given dosages above a certain minimal level.

SOME BIOCHEMICAL ASPECTS OF MENTAL DEFICIENCY

Two diseases are presented here in illustration—kernicterus and phenylpyruvic oligophrenia. Kernicterus, which is associated with mental deficiency in many cases, is the neuropsychiatric complication of erythroblastosis fetalis (9, 41). Erythroblastosis fetalis has an immunochemical basis. The most common (among over 90 per cent of patients) immunologic constellation leading to the disease is an Rh-positive father and child and an Rh-negative mother. The Rh-negative mother becomes immunized during pregnancy by the Rh antigen present in the Rh-positive fetus. As a result, the mother forms Rh antibodies which pass through the placenta to cause erythroblastosis fetalis in the child (4). About 10 per cent of children with erythroblastosis fetalis develop kernicterus (41). It cannot be predicted which of the children will develop kernicterus, but as a result of the immunologic studies—the absence of the proper immunologic constellation—it can be stated that certain children will *not* develop erythroblastosis fetalis, and therefore cannot develop its complication, kernicterus.

TABLE 2

STUDIES OF PALMAR SWEATING

SUBJECTS†	*N*	SWEATING RESPONSE* STRONG (%)	WEAK (%)
Psychiatric patients	100	84	16
Nonpsychiatric patients	1160	83	17
Hospital personnel	71	32	68

SUBJECTS‡	*N*	MEAN AMOUNT OF SWEAT (mg.)
Anxiety neurosis patients	33	19.1
Normal soldiers	54	18.0
Hospital personnel	30	9.1

* All differences shown are significantly different, $p < .05$, except between samples of anxiety neurosis patients and normal soldiers.

† Silverman and Powell (38).

‡ Cohen (7).

The other disease, phenylpyruvic oligophrenia, is a very satisfying one so far as predictive ability is concerned. Without knowing or seeing the patient, one can see that a positive chemical test for phenylpyruvic acid in the urine means that the presence of mental deficiency can be predicted in *every* such case (30). Yet, we cannot predict which unborn children will have the disease, nor can anything be said about the relation of the chemical to the brain abnormality or to the severity of mental deficiency.

STUDIES ILLUSTRATING A RELATIONSHIP BETWEEN BIOCHEMICAL FINDINGS AND PERSONAL, SUBJECTIVE FEELINGS

In the upper part of Table 2 are shown some early results in a study of palmar sweating (38) among patients in an army hospital. The sweating was measured by a nonquantitative method, which indicated that psychiatric patients and nonpsychiatric medical or surgical patients had more palmar sweating than did hospital personnel. In the lower part of Table 2 is shown another study (7) with

essentially the same results. This latter study extended the results of the earlier study in three ways: a quantitative method was used, the psychiatric patients were carefully selected with a single diagnosis, and normal soldiers rather than other patients were used as control subjects. The differences between patients and soldiers were not significant, but those between either of these groups and hospital personnel were significant. It is commonly stated that patients with anxiety neurosis have sweaty palms. However, these studies show that nonpsychiatric patients and normal soldiers have the same amount of sweatiness. On the other hand, hospital personnel did not show this phenomenon. It may be inferred, therefore, that sweaty palms do not necessarily signify anxiety neurosis, but may be related to such personal feelings as familiarity of the situation. These studies indicate the possible future value of biochemical studies in analyzing and describing subtle psychologic states.

THE USE OF RESPONSES TO AUTONOMIC DRUGS IN PREDICTING IMPROVEMENT FOLLOWING ELECTROSHOCK THERAPY (EST)

The next two studies present a more conventional kind of prediction than the preceding ones. A group of 100 patients, chiefly schizophrenic and manic-depressive, composed the experimental group (11). Before and after a course of EST they were tested on separate days with intravenous epinephrine and intramuscular mecholyl. Their blood pressure, psychologic responses, and somatic responses (pallor, sweating, and so on) were followed. The patients were classified in two different ways: (*a*) Seven different groups according to their systolic blood pressure response to mecholyl and epinephrine. Groups I to VI were classified on the basis of marked, moderate, or mild reactions to epinephrine and mecholyl. For example, Group I had a marked reaction to epinephrine (marked increase in blood pressure) and a mild reaction to mecholyl (very slightly decreased blood pressure). Group VII had a chill after the administration of mecholyl. (*b*) Four different groups according to whether anxiety was produced or relieved and whether a shaking chill after mecholyl occurred (Table 3). The favorable groups were VI, anxiety precipitated by mecholyl only, anxiety decreased by mecholyl, and chill after mecholyl (Group VII). These results are interesting and provocative; however, they have not yet been confirmed in other laboratories, nor has there been reported a follow-up study long enough to be reliable.

TABLE 3

USE OF PSYCHOLOGIC AND PHYSIOLOGIC RESPONSES TO MECHOLYL AND EPINEPHRINE TO PREDICT RESULTS OF EST*

RESPONSE TO DRUGS	*N*	IMPROVED AFTER EST (%)
1. Blood pressure changes		
a. Groups I–V	47	13
b. Group VI	34	82
2. No anxiety precipitated		6
3. Anxiety precipitated by		
a. epinephrine only	16	19
b. mecholyl only	10	100
4. Anxiety decreased by mecholyl	25	88
5. Chill after mecholyl (Group VII)	13	100

* Funkenstein, Greenblatt, and Solomon (11).

PREDICTION OF RECOVERY IN SCHIZOPHRENIA AND MANIC-DEPRESSIVE PSYCHOSIS UTILIZING THE MEASUREMENT OF BLOOD LIPIDS AND BODY BUILD

An early study of Gildea *et al.* (12) showed a significant association in normal men of a pyknic physique, high serum fatty acids, and a high energy output, and a significant association of a leptosomic physique, low serum fatty acids, and low energy output. Gildea and Man (13) then followed this with a prognostic study of schizophrenia and manic-depressive psychosis. Table 4 shows that the patient material was divided into four groups on the basis of a five year follow-up. There was a sick

TABLE 4

PREDICTION OF CAPACITY FOR RECOVERY IN SCHIZOPHRENIC AND MANIC-DEPRESSIVE PATIENTS (13)*

DIAGNOSIS	HEALTH AFTER FIVE YEARS	*N*	PYKNIC PHYSIQUE (%)	SERUM FATTY ACIDS % >12.5 meq./L	SERUM CHOLESTEROL % >200 mg.%
Schizophrenia	sick	30	10	7	3
Schizophrenia	well	22	23	55	45
Manic-depressive	sick	44	44	6	6
Manic-depressive	well	59	59	73	91

* All differences shown are significantly different, $p < .05$

and a well schizophrenic group and a sick and a well manic-depressive group at the end of this time. Within the diagnostic groups there was a significant association of pyknic physique, high serum fatty acids, and high serum cholesterol with a high capacity for recovery. This was a striking statistical phenomenon, although in an individual case it might well fail to predict capacity for recovery accurately. There were only a few exceptions to the rule that patients with high blood lipids had a high capacity for recovery.

CHEMICAL CHANGES IN THE BRAIN ITSELF

An implicit assumption underlies these biochemical studies of behavior—behavior, subjective or otherwise, is associated with biochemical changes in the brain. So far, the chief evidence that this is a fact rather than only an assumption is contained in a series of papers (8, 31, 32, 33) published since 1948 (Table 5). These papers present a series of experiments performed on rats which were "emotionally excited" by inducing repetitive loss of support for about four minutes. The *whole* brain was then analyzed for the chemical constituents listed in the table. There were significant changes in three components: lactic acid, acetyl choline, and phosphocreatine. The other components did not show an appreciable change. These changes do not seem to form any coherent pattern that allows unequivocal inferences concerning the mechanisms involved in the brain during emotional excitement. One difficulty in this kind of study is that the whole brain must be used for analysis. Since the brain is so complex anatomically, the use of the whole brain may obscure crucial changes within a small group of cells, and consequently makes the interpretation of any changes found uncertain.

One approach used to analyze smaller pieces of brain tissue is that initiated by Caspersson (5) in Sweden, and applied to psychiatric problems by Hydén (16). This method achieves ultra-violet absorption spectrophotometry of individual cells. By means of a complex microscopal setup ultra-violet light is thrown on various parts of the cell. The resulting absorption spectrum results from the presence of proteins and nucleic acids in the cell. Hydén (16) reported significant changes in the absorption spectrum of nerve cells in schizophrenia.

TABLE 5

EFFECTS OF EMOTIONAL EXCITEMENT ON VARIOUS CHEMICAL CONSTITUENTS OF WHOLE RAT BRAINS

CHEMICAL CONSTITUENT	NORMAL VALUE mg.%	EXCITEMENT VALUE mg.%
Lactic acid (33)	18.8	37.5*
Acetyl choline (μg/gm) (31)	1.25	0.87*
Phosphocreatine (8)	9.8	11.1*
Ammonia (32)	0.28	0.22
Glutamine (32)	79.0	78.0
Inorganic phosphate (8)	13.8	14.0
Adenosine triphosphate (8)	14.9	14.9
Hexose phosphate (8)	16.8	16.7
Total phosphate (8)	55.3	56.7

* These differences are significant, $p < .05$.

Although this work has been criticized by Meyer (27) and MacKinnon (24), this, of course, does not mean that the cytochemical method as such cannot be applied to psychiatric problems.

The chief difficulty of this spectrophotometric method of Caspersson is its limitation to the measurement of only proteins, some lipids, and nucleic acids. Another approach to the examination of small pieces of brain tissue is that initiated by Linderstrøm-Lang (22) for the stomach. These methods are being elaborated for the brain in the Department of Pharmacology, Washington University School of Medicine,[2] and in the Department of Neuropsychiatry (34, 35). The techniques depend upon the possibility of analyzing minute bits (micrograms) of brain tissue for chemical constituents, including enzymes. Because only minute amounts of tissue are necessary, a relatively precisely defined part of the brain may be studied, and, therefore, the objection of working on heterogeneous areas of the brain is overcome. For example, the layers of the cerebellum (34) can be analyzed separately so that it is possible not only to compare one layer with another, but individual layers of this region can be compared with histologically similar layers of other areas of the brain. It is our hope eventually to apply these methods to the study of human psychiatric disease.

SOME METHODOLOGIC CONSIDERATIONS INVOLVED IN PREDICTING BEHAVIOR

Biochemical tests generally give precise, quantitative, and reproducible results. The use

[2] Lowry, O. H., *et al.* Unpublished data.

of such valid and quantitative tests in investigating human behavior does not lessen the need for careful, controlled observation of the behavior. Since few biochemists are trained in the observing and interpreting of human behavior, the importance of this aspect of the study is often overlooked by them. This idea that the use of a biochemical test in a study automatically insures the excellence of the study is held not only by biochemists but also by psychologists and psychiatrists. For this reason, certain methodologic problems should be emphasized. Without attention to these problems, the interpretation of the biochemical test may become meaningless in terms of its relation to human behavior.

The first problem arises in selecting the experimental group. The group must be selected by criteria defined in such a fashion that they are precisely delineated, understandable, and reproducible by other workers who may desire to select a similar group in order to repeat or extend the experiments. The important issue is that in speaking of diagnosis, psychologic state, physiologic state, or socioeconomic state, we define them in an unequivocal fashion that is so clear that other workers can read what we have done and reproduce the work. This means we must sacrifice a certain intuitive subtlety and study a "nuclear" group, i.e., one that fits our definition unequivocally, even though this excludes borderline cases who will some day be known to belong to the same group when an objective, quantitative method of selection is discovered.

The second problem arises in the selection of criteria for classifying the predicted behavior. This has already been mentioned above in the discussion of low blood sugar. There we contrasted our excellent ability to predict the occurrence of a confusional psychosis with our much inferior ability to predict precisely what the behavior of a given patient may be while confused because of a low blood sugar. We must define our criteria quite carefully, therefore; otherwise the data obtained can be distorted to support or attack any idea we may like or dislike at the moment. As a part of the problem of the definition of the predicted behavior, considerations about the significant aspects of the behavior in time are relevant. A follow-up study on such serious and chronic diseases as schizophrenia or manic-depressive psychosis that follows the patients only one month or even six months may be of limited value, but more probably is even worse than being of limited value—it may actually be misleading. In the case of psychiatric disease and psychologic behavior of normal subjects, so little is known about the natural history of psychiatric disease or of normal behavior that it is often difficult to decide how long a follow-up is indicated. In this circumstance it seems wiser to use much longer follow-up studies than are now used in most instances.

The third problem concerns the predictive technique utilized. This requires the use of a technique that gives unequivocal, consistent results which do not require special interpretation in the light of the predictive study being done. This, it would seem, would preclude the use of uncontrolled clinical observation, and would preclude the use of complicated tests that have not been carefully validated.

The fourth consideration involves the method used in collecting the data about behavior. The chief point of emphasis is that poorly planned experiments and poorly collected data do not lend themselves to precise interpretation or to valid statistical analysis. For example, when the statement that a certain per cent of a given group show "hostility" is made, unless there was a standard reproducible way of deciding whether "hostility" was present or not, and unless this method was used routinely on every patient, the calculation of a percentage is misleading. Finally, there is the area (2, 28) of the often omitted but crucial factors involved in cross validation and validity generalization. It is this application of an apparently substantiated predictive hypothesis to additional samples of the research population which provides the most rigorous test of the validity of our predictive hypothesis.

SUMMARY

A series of illustrations of predictive methods reviews some of the techniques of biochemistry and physiology in their application to psychological and psychiatric problems. The contributions of these interrelationships are indicated, together with methodologic problems of experimental control.

REFERENCES

1. ADAMS, R. D., & KUBIK, C. S. Subacute degeneration of the brain in pernicious anemia. *N. Eng. J. Med.*, 1944, **231**, 1–9.
2. AULD, F., JR., & ERON, L. D. The use of Rorschach scores to predict whether patients will continue psychotherapy. *J. consult. Psychol.*, 1953, **17**, 104–109.
3. BERLE, B. B., PINSKY, R. WOLF, S., & WOLFF, H. G. A clinical guide to prognosis in stress diseases. *J. Amer. med. Ass.*, 1952, **149**, 1624–1628.
4. CAPPELL, D. F. The mother-child incompatibility problem in relation to nervous sequelae of haemolytic disease of the newborn. *Brain*, 1947, **70**, 486–494.
5. CASPERSSON, T. O. *Cell growth and cell function: a cytochemical study.* New York: Norton, 1950.
6. CLECKLEY, H. M., SYDENSTRICKER, V. P., & GEESLIN, L. E. Nicotinic acid in the treatment of atypical psychotic states associated with malnutrition. *J. Amer. med. Ass.*, 1939, **112**, 2107–2110.
7. COHEN, M. E. Studies of palmar hand sweat in healthy subjects and in patients with neurocirculatory asthenia (anxiety neurosis, neurasthenia, effort syndrome) with a description of a simple quantitative method. *Amer. J. med. Sci.*, 1950, **220**, 496–507.
8. DAWSON, R. M. C., & RICHTER, D. Effect of stimulation on the phosphate esters of the brain. *Amer. J. Physiol.*, 1950, **160**, 203–211.
9. EVANS, P. R., & POLANI, P. E. The neurological sequelae of Rh sensitization. *Quart. J. Med.*, 1950, **19**, 129–149.
10. FELDMAN, M. J. A prognosis scale for shock therapy. *Amer. Psychologist*, 1948, **3**, 348. (Abstract)
11. FUNKENSTEIN, D. H., GREENBLATT, M., & SOLOMON, H. C. A test which predicts the clinical effects of electric shock treatment on schizophrenic patients. *Amer. J. Psychiat.*, 1950, **106**, 889–901.
12. GILDEA, E. F., KAHN, E., & MAN, E. B. The relationship between body build and serum lipoids and a discussion of these qualities as pyknophilic and leptophilic factors in the structure of the personality. *Amer. J. Psychiat.*, 1936, **92**, 1247–1260.
13. GILDEA, E. F., & MAN, E. B. Methods of estimating recovery in patients with manic-depressive and schizophrenic psychoses. *Amer. J. Psychiat.*, 1943, **99**, 496–506.
14. GOLDBERG, L. Quantitative studies on alcohol tolerance in man. *Acta physiol. Scand.*, 1943, **5**, Suppl. 16, 1–128.
15. HORST, P. (Ed.) The prediction of personal adjustment. *Soc. Sci. Res. Coun. Bull.*, 1941, No. 48.
16. HYDÉN, H., & HARTELIUS, H. Stimulation of the nucleoprotein-production in nerve cells by malonitrile and its effect on psychic functions in mental disorders. *Acta psychiat., Kbh.*, 1948, Suppl. 48, 1–117.
17. JOLLIFFE, N., GOODHART, R., GENNIS, J., & CLINE, J. K. The experimental production of vitamin B_1 deficiency in normal subjects: the dependence of urinary excretion of thiamin on the dietary intake of vitamin B_1. *Amer. J. med. Sci.*, 1939, **198**, 198–211.
18. JOLLIFFE, N., WORTIS, H., & STEIN, M. H. Alcoholic encephalopathies and nutrition. In E. M. Jellinek (Ed.), *Alcohol addiction and chronic alcoholism.* New Haven: Yale Univer. Press, 1942.
19. KLINE, N. S., & TENNEY, A. M. Constitutional factors in the prognosis of schizophrenia. *Amer. J. Psychiat.*, 1950, **43**, 240–258.
20. KRANTZ, J. C., JR., & CARR, C. J. *The pharmacologic principles of medical practice.* Baltimore: Williams & Wilkins, 1949.
21. LEVIN, M. Experimental bromide intoxication: a criticism. *Amer. J. Psychiat.*, 1951, **108**, 309–310.
22. LINDERSTRØM-LANG, K., HOLTER, H., & OHLSEN, A. S. Studies in enzymatic histochemistry: XIII. The distribution of enzymes in the stomach of pigs as a function of its histological structure. *Compt. rend. Trav. Lab. Carlsberg, Serie Chim.*, 1935, **20**, 66–125.
23. LIPMANN, F. Biosynthetic mechanisms. *Harvey Lect.*, 1948, **24**, 99–123.
24. MACKINNON, I. H., HOCH, P. H., CAMMER, L., & WAELSCH, H. B. The use of malonitrile in the treatment of mental illness: preliminary report. *Amer. J. Psychiat.*, 1949, **105**, 686–688.
25. MALAMUD, D. I. Objective measurement of clinical status in psychopathological research. *Psychol. Bull.*, 1946, **43**, 240–258.
26. MENSH, I. N. Statistical techniques in present-day psychodiagnostics. *Psychol. Bull.*, 1950, **47**, 475–492.
27. MEYER, A., & MEYER, M. Nucleoprotein in nerve cells of mental patients: a critical remark. *J. ment. Sci.*, 1949, **95**, 180–181.
28. MOSIER, C. I. Problems and designs of cross-validation. *Educ. psychol. Measmt*, 1951, **11**, 5–11.
29. MYERS, C. R. Prediction in clinical psychology. *Canad. J. Psychol.*, 1950, **4**, 97–108.
30. PENROSE, L. S. *The biology of mental defect.* New York: Grune & Stratton, 1949.
31. RICHTER, D., & CROSSLAND, J. Variation in acetyl choline content of the brain with physiological state. *Amer. J. Physiol.*, 1949, **159**, 247–255.
32. RICHTER, D., & DAWSON, R. M. C. Brain metabolism in emotional excitement and in sleep. *Amer. J. Physiol.*, 1948, **154**, 73–79.
33. RICHTER, D., & DAWSON, R. M. C. The ammonia and glutamine content of the brain. *J. biol. Chem.*, 1948, **176**, 1199–1210.
34. ROBINS, E., & SMITH, D. E. A quantitative histochemical study of eight enzymes of the cerebellar cortex and subjacent white matter. *Proc. Ass. Res. nerv. ment. Dis.*, 1953, **32**, 305–327.
35. ROBINS, E., SMITH, D. E., & MCCAMAN, R. E. Microdetermination of purine nucleoside phosphorylase activity in brain and its distribution within the monkey cerebellum. *J. biol. Chem.*, 1953, **204**, 927–937.

36. ROSENZWEIG, S. A method of validation by successive clinical predictions. *J. abnorm. soc. Psychol.*, 1950, **45**, 507–509.
37. SARBIN, T. R. The logic of prediction in psychology. *Psychol. Rev.*, 1944, **51**, 210–228.
38. SILVERMAN, J. J., & POWELL, V. E. Studies on palmar sweating: III. Palmar sweating in an army general hospital. *Psychosom. Med.*, 1944, **6**, 243–249.
39. STEIN, M., RONZONI, E., & GILDEA, E. F. Physiologic response to heat stress and ACTH of normal and schizophrenic subjects. *Amer. J. Psychiat.*, 1951, **108**, 450–455.
40. STOCKINGS, G. T. A clinical study of the mescaline psychosis, with special reference to the mechanism of the genesis of schizophrenic and other psychotic states. *J. ment. Sci.*, 1940, **86**, 29–47.
41. VAUGHAN, V. C., III, ALLEN, F. H., JR., & DIAMOND, L. K. Erythroblastosis fetalis: IV. Further observations on kernicterus. *Pediatrics*, 1950, **6**, 706–716.
42. WILLIAMS, R. D., & MASON, H. C. Further observations on induced thiamine (vitamin B_1) deficiency and thiamine requirement of man. *Proc. Staff Meet. Mayo Clin.*, 1941, **16**, 433–438.
43. WILLIAMS, R. D., MASON, H. L., WILDER, R. M., & SMITH, B. F. Observations on induced thiamine (vitamin B_1) deficiency in man. *Arch. intern. Med.*, 1940, **66**, 785–799.
44. WITTMAN, PHYLLIS. Diagnostic and prognostic significance of the shut-in personality type as a prodromal factor in schizophrenia. *J. clin. Psychol.*, 1948, **4**, 211–214.
45. ZUBIN, J. Introduction: symposium on statistics for the clinician. *J. clin. Psychol.*, 1950, **6**, 1–6.

Received February 21, 1953.

AFFECTIVE TONE AND VISUAL RECOGNITION THRESHOLDS

LEONARD D. GOODSTEIN

State University of Iowa

McGINNIES (5) and others (12) have invoked a concept of perceptual defense to account for the heightened recognition thresholds for negatively valued and affectively charged words. Perceptual defense supposedly operates as a mechanism of resistance to the recognition of such anxiety-provoking stimuli.

McGinnies' interpretation, however, has been severely criticized by Howes and Solomon (2) on the grounds that it did not take into account the relative frequency of the stimulus words as they occur in printed material. These critics pointed out that the heightened thresholds of the words with negative affective value can be more parsimoniously explained in terms of their associative strength or relative familiarity. Additional evidence of the importance of the associative value of the stimulus words was found in several subsequent studies (3, 5, 6, 7, 8, 9, 10).

The purpose of the present note is to demonstrate that the recently reported results of Aronfreed, Messick, and Diggory (1) can also be interpreted as a function of the relative associative strength of the stimulus words. These investigators studied tachistoscopic visual recognition thresholds and galvanic skin responses as functions of the affective value of stimulus words. The 25 words employed had previously been scaled from pleasant to unpleasant by the method of equal-appearing intervals. The words were placed in three categories: pleasant ($N = 5$), neutral ($N = 5$), and unpleasant ($N = 15$). The authors reported that the mean visual recognition threshold for the unpleasant words was significantly higher ($p < .01$) than the means for the neutral and the pleasant words. There were no significant differences in recognition thresholds between the pleasant and neutral words.[1] The authors interpreted their threshold data as confirming McGinnies' (5) previous finding that "the affective tone of the stimulus words has had an effect on mean recognition time . . . (1, p. 525).

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR THE LOG WORD FREQUENCIES OF THREE CATEGORIES OF STIMULUS WORDS

STATISTIC	GROUP		
	PLEASANT	NEUTRAL	UNPLEASANT
N	5	5	15
Mean	2.85	2.63	1.52
SD	.83	.52	.50

Examination of their list revealed that, of the ten pleasant and neutral words, five were among the 500 most common words in the English language as determined by the Thorndike-Lorge word count (11); seven of the ten were among the 1,000 most common words. Only one of the 15 unpleasant words, however, was among the 1,000 most common words.

The relative frequency of each of the 25 words was determined by use of the Thorndike-Lorge general word count (11). The means and standard deviations of the logs of these frequencies for the three categories of words are presented in Table 1. The differences between the mean log frequencies of the unpleasant and neutral words ($t = 3.53$, $df = 18$, $p < .01$) and between the unpleasant and pleasant words ($t = 2.68$, $df = 18$, $p < .02$) are highly significant. There is no difference between the means of the neutral and pleasant words ($t = .39$, $df = 8$, $p < .60$).

The correlation between the pleasantness ratings of the stimulus words and their recognition thresholds is .53 ($p < .01$).[2] When, however, the variance attributable to the associative value (relative frequency) is partialled out, this correlation is reduced to .18 ($p = .34$ approx.).

The associative value of the stimulus words is not the only factor that can modify a

[1] The GSR data are difficult to interpret because of a sex and instructional set interaction, and will not be considered in the present paper.

[2] The writer wishes to thank Dr. James Diggory for making available additional data for this analysis that were not reported in the published paper.

perceptual defense interpretation. The effects of response suppression, i. e., the voluntary refusal to report emotionally charged stimulus words, have also been clearly shown (7). Aronfreed, Messick, and Diggory attempted to study the effects of response suppression by giving one group of *S*s information about the nature of the words that would be used as stimuli. The correlation between the mean log frequencies and recognition thresholds for the group receiving this information is not significantly different from that for the non-informed group. The correlations were −.62 and −.54, respectively. This finding would strongly suggest that the authors' attempts to prevent response suppression did not reduce the importance of the associative value of the stimulus words.

These findings clearly confirm the results of those studies that have found no differences in recognition thresholds for stimulus words of purportedly differential affective value when the relative frequency of the stimulus words is controlled (7, 8, 9).

REFERENCES

1. ARONFREED, J. M., MESSICK, S. A., & DIGGORY, J. C. Re-examining emotionality and perceptual defense. *J. Pers.*, 1953, **21**, 517–528.
2. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
3. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
4. MCGINNIES, E. M. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
5. MCGINNIES, E. M., COMER, P. B., & LACEY, O. L. Visual recognition thresholds as a function of word length and word frequencies. *J. exp. Psychol.*, 1952, **44**, 65–69.
6. MURPHY, D. B. Recognition and recall as a function of frequency, shock and individual differences variables. Unpublished doctor's dissertation, State Univer. of Iowa, 1953.
7. POSTMAN, L., BRONSON, WANDA S., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–224.
8. POSTMAN, L., & SCHNEIDER, B. M. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, **58**, 271–284.
9. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.
10. SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195–201.
11. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teachers Coll., Columbia Univer., 1944.
12. VANDERPLAS, J. M., & BLAKE, R. R. Selective sensitization in auditory perception. *J. Pers.*, 1949, **18**, 252–256.

Received February 15, 1954. Prior publication.

SOME EFFECTS OF SITUATIONAL THREAT ON GROUP BEHAVIOR[1]

JOHN T. LANZETTA

Crew Research Laboratory, Air Force Personnel and Training Research Center

DON HAEFNER, PETER LANGHAM, AND HOWARD AXELROD

University of Rochester

THE effects of threat, or other anxiety-evoking situations, on behavior have received increasing attention from social scientists in recent years. Interest has been evidenced both by an increase in empirical studies (5, 7, 8, 9, 10) and renewed attempts to incorporate the concept of anxiety within the framework of contemporary learning theory (13, 15, 18). The majority of the experiments have concerned themselves with the effects of situationally produced anxiety on performance and have focused on individuals either acting alone or coacting. Such studies of necessity ignore the area of interpersonal relations, an area which clinical and theoretical literature would indicate is particularly sensitive to changes in anxiety level. The present study was an attempt to determine the effects of situational threat on the behavior of *interacting individuals*. As such it can contribute to the growing body of empirical evidence on the functioning of small groups as well as shed light on the experimentally neglected area of anxiety and interpersonal relations.

The broad question to be examined is this: How does variation in threat to a group reflect itself in changes in group member behavior? Variation in threat can occur along several dimensions: nature and/or intensity, locus in time, locus in "psychological" space, and target. The latter two are peculiar to a concern with group processes and have no counterpart if the focus were the individual acting alone. "Locus in psychological space"[2] refers to the origin of the threat: whether it arises from some *internal* group process or originates from an *external-to-the-group* source. Target is orthogonal to this dimension; either the threat is directed to the group as a whole or to individual group members. These two dichotomized dimensions were the independent variables for this study.

[1] The work described in this paper was done under a contract between the Office of Naval Research and the University of Rochester.

[2] In some respects this dimension when limited to the dichotomy "no-locus" (objectless, unspecified, etc.) and "specific-locus" parallels the distinction made between anxiety and fear. We are interested only in situationally produced threat and thus according to some theorists are dealing with fear. We concur with the opinion of Miller and Dollard (14) and will not attempt to distinguish between fear and anxiety.

METHOD

The basic method used was that reported in previous studies by this laboratory (2, 3). Groups composed of three subjects (*Ss*) worked on a task called "intercept" in a laboratory setting. The *Ss* were observed by four observers (*Os*), who immediately classified their ongoing behavior in terms of a category system. After the experimental session, each *S* was rated on a number of traits and both *Os* and *Ss* filled out a questionnaire designed to assess various "group" characteristics. The *Ss* also responded to an adjective check list and a social-sensitivity questionnaire.

Design. The main independent variable was threat. Within the threat condition, a distinction was made between locus and target of threat, each of these being dichotomized: locus into external and internal to the group, and target into group or individually directed. A 2 × 2 factorial design was utilized for the threat conditions with nine groups assigned to each of the cells. Nine groups were also assigned to a control (nonthreat) condition.

Experimental conditions. Threat was produced by structuring the sessions as evaluation tests. Previous studies of leadership at the laboratory have utilized ROTC trainees, and subsequent inquiries indicated that they perceived the project staff as an evaluative group. By securing the cooperation of the AFROTC and NROTC commanding officers, we could utilize this orientation to create our threat. The *Ss* were informed by letter that they were to be evaluated and that this assessment would become part of their record. This was reinforced by several impersonal letters, rather curt notices as to scheduled sessions, and a formal and impersonal attitude on the part of the experimenter (*E*).

Within this context of external threat, target and locus were varied by varying instructions as to the method of evaluation. Locus was manipulated by varying the stated source of the evaluation ratings with either *Os* rating *Ss* (external threat), or *Ss* rating themselves (internal threat). Target was manipulated by informing *Ss* that they were being evaluated as individuals (individual threat), or that their evaluation reports would be identical and based on ratings the group received (group threat).

After the instructions were read, *S*s were told to proceed to the task instructions, and *E* left the room. He did not re-enter the room again until completion of the task, at which time *S*s were asked to fill out some questionnaires and then allowed to go.

The *S*s under control conditions were instructed by a different *E*. They were informed that they were being used for a study necessary to complete requirements for a graduate course. The *E* attempted to set an informal, relaxed atmosphere when *S*s first entered and usually spent five to ten minutes interacting with them before the session began. The *S*s were asked to do their best, the task instructions were pointed out, and they were told to proceed.

Subjects. The *S*s for the threat condition were 108 AFROTC and NROTC students in freshman and sophomore classes. They comprised about 60 per cent of the total enrollment of these groups. The 108 *S*s were randomly assigned to groups of three, and the groups were then randomly assigned to cells within the threat condition.

The 27 *S*s for the control condition were volunteers from the freshman and sophomore non-ROTC student population. They were randomly assigned to nine groups of three members each.

Task. The task utilized was called "intercept" and was a modification of a task developed by Schutz[3] for the Tufts Systems Analysis project. The essential elements were a 4 × 4-foot board ruled into 11 × 11 squares each identified by coordinates, and enemy and friendly planes of three different colors.[4] The group's job was to protect a "carrier" (the center square on the board) from being hit by "enemy" planes. The members accordingly had to plan a strategy for deployment of their planes against the enemy attack. Since but one of their planes could move in response to each enemy move, they were required to decide jointly on the most effective move to make. Several factors were introduced to make such choices a possible source of conflict: the scoring was so arranged that there were penalties and rewards for the group as a whole as well as for individual members of the group; a different suggested "best" strategy was imbedded in each *S*'s instructions; the "radar" reports for each member were usually different. There were many occasions in the action where the individual was presented the alternative of behaving in his own or the group interest.

The full task consisted of three phases. In Phase 1 *S*s worked on the intercept problem for 15 minutes. Phase 2 was a 10-minute discussion of a "best" strategy based on *S*s' experience in Phase 1. In Phase 3 they again worked for 15 minutes on the intercept problem (a new set of enemy moves), utilizing the strategy decided upon in Phase 2.

All instructions for the task were typed, and those for the various phases were placed in separate envelopes. The *S*s were informed when to move to a new phase by a buzzer signal. The *E* was not present at any time after the reading of the initial instructions discussed in the preceding section.

Instruments. The primary data-collection technique was the recording of ongoing behavior in terms of an adaptation of the system proposed by Bales (1). The system was elaborated by the use of subcategories used by Bales as significata for his main categories. A full description of the categories is presented by Bales (1, pp. 177–195).

One possible criticism of an analysis based on categorizations of ongoing behavior is that indices based on a simple sum of behaviors in certain categories ignore the differential intensity of similarly classified acts and the integrated, holistic character of an individual's behavior. Since to a certain extent this criticism may be valid, global judgments were made of behavioral characteristics of individuals-in-situation by means of rating scales. These ratings consisted of judgments about the cooperativeness, initiative, effective leadership, goal orientation, etc. of individuals. In some cases the ratings paralleled indices derived from the interaction categories, and thus served as a check on them.

Another area in which global judgments seemed necessary was that of group properties. Properties such as the morale or cohesiveness of a group are probably the synthesis of certain interaction patterns; but, because concise behavioral definitions of the terms are lacking, it was difficult to arrive at an index based on the categories alone. For this purpose, a postmeeting reaction questionnaire was developed. This questionnaire, with the addition of certain satisfaction and sociometric items, was also filled out by *S*s. The *S*s were also asked to check an adjective check list[5] and a social-sensitivity inventory.[6]

The number of dependent variables measured by use of these instruments can be classified into three broad areas. The rational classes, which crosscut the instruments, are:

1. Variables attempting to assess the nature of the interpersonal relations between group members. The *denotata* of this class are aggression, sociability, cooperativeness, autocratic behavior, solidary (friendly, praising, etc.) behavior, antagonistic acts, striving for group acceptance, the degree of informal friendliness exhibited by the group, and the extent of personality conflict.

2. Variables concerned with utilization of the actors' and other group members' resources in attempted problem solution. These are sometimes difficult to distinguish from 1 and in fact are not considered completely independent. The differentiating criterion is whether problem solution or another group member is considered the primary target of action. Included in this class are such variables as leadership; self-oriented behavior; achievement motivation; giving of suggestions, opinions, and orientations; asking for opinions, suggestions, or orientations; disagreements and agreements; goal

[3] We gratefully acknowledge the aid of Dr. W. C. Schutz in making the task available to us.

[4] A complete description of the task is unfeasible in the short space available. A fuller description and instructions to *S*s can be obtained from the authors.

[5] Developed by Helen and Vincent Nowlis at the University of Rochester, from a list utilized by Gough at the University of California.

[6] Developed by Urie Bronfenbrenner at Cornell University.

orientation of the group; competition in the group; and the degree of equal participation of group members in problem solution.

3. Variables that attempt to assess the effectiveness of the group or the individual in achieving specified ends. Adaptability, efficiency, productivity, and communication effectiveness in the group are included in this class.

The first two classes require inferences about motivational states from observation of ongoing behavior and, as mentioned, the distinguishing criterion is the target of action. The last class requires judgments about the extent to which behavior resulted in certain specified effects.

RESULTS

Measures obtained from records of ongoing behavior, ratings of each individual, ratings of the group, *S*s' responses to a questionnaire concerned with aspects of their group's functioning, and *S*s' responses to an adjective check list were available for analysis. For present purposes a "group" score for each variable was obtained by summing over observers. Where measures on individuals were obtained, scores were also summed over the three individuals composing a group. For the Bales category system, the group score on each category of behavior was then expressed as a proportion of the total number of acts recorded for the group. Such a proportion index eliminates the possibility of obtaining significant differences between treatments due to a difference in activity rate. Its use was based on the assumption that change in activity rate would be proportionately distributed over the categories in the system.

TABLE 1

INTEROBSERVER RELIABILITIES FOR THE BALES CATEGORIES

(N = 108)

CATEGORY	r_{tt}*
I Shows solidarity	.64
II Shows tension release	.65
III Agrees	.83
IV Gives suggestion	.89
V Gives opinion	.91
VI Gives orientation	.87
VII Asks orientation	.72
VIII Asks opinion	.80
IX Asks suggestion	.46
X Disagrees	.76
XI Shows tension	.70
XII Shows antagonism	.71

* Corrected for double length since the sum of the two *O*s' categorizations were used in analysis.

TABLE 2

INTEROBSERVER RELIABILITIES FOR TRAIT RATINGS

(N = 108)

RATING	AVERAGE RELIABILITY OF SINGLE OBSERVER (FORM)*
1. Cooperativeness	.46
2. Efficiency	.66
3. Leadership	.71
4. Aggression	.53
5. Autocratism	.62
6. Sociability	.38
7. Striving for group acceptance	.34
8. Adaptability	.55
9. Self-oriented needs	.43
10. Achievement motivation	.65

* See text for discussion of technique used in obtaining these reliabilities.

The interobserver reliabilities for the Bales categories are shown in Table 1. The interobserver reliabilities for the subcategories were considerably lower than those for the main categories, and thus only the main categories were used for analysis.

The average observer reliabilities of the trait ratings are shown in Table 2. The best technique to use in determining multijudge reliability is difficult to decide upon (6, 17), and estimates obtained by various methods are not exactly comparable. The reliabilities shown were obtained by an analysis of variance technique described by McNemar (11, p. 280) and are based on an error of measurement estimate which includes an adjustment for differing observer (form) means. They are underestimates of the "true" reliabilities since scores utilized for analysis were based on ratings summed over four *O*s as well as three *S*s.

In general, results will be presented in terms of measuring instruments and then organized with respect to the areas of functioning discussed previously.

As the statement of the problem indicated, the study was concerned with two main questions: (*a*) What are the effects of situational threat on certain aspects of group functioning? (*b*) To what extent do locus and target of threat affect group member behavior?

Although the former question has logical priority, the design dictated an analysis for locus and target as a first step.

Locus and Target of Threat

Table 3 indicates the mean percentage of behavior falling in the various Bales categories for the four experimental conditions. The percentages are believed to be quite stable, the number of units on which they are based ranging from 1,100 to 1,300.

From the total column one can note considerably more of some kinds of behavior recorded than others. Most of this difference reflects the nature of the task and the composition of the group, but some of it probably results from the differential difficulty of observing various kinds of behavior. The distribution obtained is similar to one obtained by Bales for a thesis discussion group and has a family resemblance to those obtained for his standard chess task (1).

The *F* ratios were computed for main effects and interaction effects for each category, by use of analysis of variance techniques. None of the *F* ratios for any category on either instrument was significant. The Bartlett test for homogeneity of variance indicated significant corrected chi squares for four of the categories.

Similar analysis for observer trait ratings, observer ratings of the group, and *S*s' ratings of their groups also yielded nonsignificant *F* ratios for main effects and interaction on all of the variables. For these instruments three of the corrected chi squares were significant. The general lack of significance for locus and target did not warrant transformations for those few variables that were nonhomogeneous since no justifiable conclusions could be based on these variables alone.

TABLE 3

MEAN PERCENTAGE OF BEHAVIOR IN THE BALES CATEGORIES FOR THE FOUR EXPERIMENTAL CONDITIONS

CATEGORY NUMBER	EXTERNAL		INTERNAL		MEAN PERCENTAGE
	A INDIVIDUAL	B GROUP	C INDIVIDUAL	D GROUP	
I Shows solidarity	2.04	1.21	1.72	1.56	1.63
II Shows tension release	.88	.60	.80	.71	.75
III Agrees	6.67	6.50	6.73	7.24	6.79
IV Gives suggestions	9.21	9.16	7.59	9.59	8.89
V Gives opinions	37.74	37.24	37.61	34.72	36.83
VI Gives orientation	22.30	23.84	24.76	25.74	24.16
VII Asks for orientation	5.23	5.41	5.24	5.42	5.33
VIII Asks for opinion	4.62	5.08	4.87	4.94	4.88
IX Asks for suggestion	.82	.97	.68	.89	.84
X Disagrees	4.09	4.26	3.84	2.63	3.71
XI Shows tension	3.78	3.17	3.43	4.30	3.67
XII Shows antagonism	2.61	2.62	2.69	2.24	2.54
Total percentage	100.00	100.06	99.96	99.98	100.02
Total number of acts	1249	1247	1277	1176	1237

Threat versus Control (Nonthreat)

Since there were no significant differences between the various threat treatments, all threat groups were combined for comparison with the control groups. The *t*'s for all of the variables were computed, using the technique described in McNemar (11, p. 223) for samples with unequal *N*'s. The results of the comparisons for the interaction categories, trait ratings, and observer ratings of the group are shown in Tables 4, 5, and 6. In those cases where the *F* ratio indicated a significant difference in variance between the two samples, Cochran and Cox's correction, as given in Edwards (4,

TABLE 4

SIGNIFICANCE OF DIFFERENCES BETWEEN THREAT AND CONTROL CONDITIONS IN MEAN PERCENTAGE OF BEHAVIOR IN THE BALES CATEGORIES*

CATEGORY	THREAT (N = 36)		CONTROL (N = 9)		F†	t†
	M	σ^2	M	σ^2		
I Solidarity	1.63	.85	1.89	.52	1.61	<1
II Tension release	.77	.32	.74	.51	−1.58	<1
III Agrees	6.79	3.18	7.24	2.80	1.14	<1
IV Gives suggestions	8.89	4.28	11.04	2.57	1.67	−3.39§
V Gives opinion	36.83	22.06	37.82	16.16	1.37	<1
VI Gives orientation	24.16	15.69	22.74	10.24	1.14	<1
VII Asks orientation	5.33	2.43	4.62	1.15	2.12	5.08§
VIII Asks opinions	4.89	2.43	4.06	2.03	1.20	1.52
IX Asks suggestions	.84	.21	.89	.20	1.05	<1
X Disagrees	3.71	4.38	2.81	2.32	1.89	1.45
XI Shows tension	3.67	2.38	3.20	2.33	1.00	<1
XII Shows antagonism	2.54	3.05	2.92	4.41	−1.45	<1
Total activity	1237	23,950	1392	9152	2.62	−3.77§

* Number of acts in a category were summed over *O*s and *S*s in a group to give a group score. These scores were then expressed as percentage of total acts recorded for the group.

† Positive values of *F* and *t* indicate threat > control; negative values indicate control > threat.

§ Significant at the .01 level or beyond.

TABLE 5

SIGNIFICANCE OF DIFFERENCES BETWEEN THREAT AND CONTROL CONDITIONS IN MEAN RATINGS OF INDIVIDUALS*

RATING	THREAT ($N = 36$)		CONTROL ($N = 9$)		F†	t†
	M	σ^2	M	σ^2		
1. Cooperation	55.78	63.38	53.22	46.94	1.35	<1
2. Efficiency	51.75	62.42	56.00	14.75	4.23‡	−2.31‡
3. Leadership	49.56	41.34	52.11	8.61	4.80‡	−1.76
4. Aggression	35.81	116.16	43.00	153.00	−1.32	−1.60
5. Autocratism	39.86	80.01	47.00	108.00	−1.35	−1.88
6. Sociability	44.92	92.02	43.56	105.02	−1.14	<1
7. Striving for group acceptance	43.72	44.03	40.00	10.25	4.30‡	2.42‡
8. Adaptability	54.08	63.51	57.89	2.36	26.91§	−2.68‡
9. Self-oriented needs	42.03	92.83	41.67	69.50	1.34	<1
10. Achievement motivation	59.33	65.71	64.56	54.54	1.21	−1.86

* Ratings were summed over the four *O*s and three *S*s in each group to give a group rating.

† Positive values of F and t indicate threat > control; negative values indicate control > threat.

‡ Significant at .05 level or beyond.

§ Significant at .01 level or beyond.

TABLE 6

SIGNIFICANCE OF DIFFERENCES BETWEEN THREAT AND CONTROL CONDITIONS IN OBSERVER RATINGS OF GROUP CHARACTERISTICS*

INDEX	THREAT ($N = 36$)		CONTROL ($N = 9$)		F†	t†
	M	σ^2	M	σ^2		
1. Goal orientedness	91.36	211.84	93.67	195.75	1.08	<1
2. Productivity	13.89	115.30	14.67	5.00	23.06§	<1
3. Degree of equal participation	39.61	100.07	42.44	120.53	−1.20	<1
4. Degree of personality conflict	44.58	231.45	46.33	219.25	1.06	<1
5. Amount of informal friendliness	51.44	167.97	51.00	277.25	−1.65	<1
6. Communication effectiveness	48.81	189.96	52.67	41.50	4.51‡	−1.23
7. Competition between members	43.61	207.44	48.33	195.50	1.06	<1
8. Cohesiveness	64.17	259.00	64.00	148.75	1.74	<1

* The index for each characteristic consisted of from three to six items, with a mean of 4.1 items per index. Each item was rated on a five-point scale and the item scores summed to give an index score. Ratings by four *O*s were summed to give a group score on each index.

† Positive values of F and t indicate threat > control; negative values indicate control > threat.

‡ Significant at .05 level or beyond.

§ Significant at .01 level or beyond.

p. 168), was used in evaluating the significance of the obtained t.

Of the 31 comparisons shown in the three tables, two were significant at the .01 level or beyond, three were significant between the .01 and .05 level, and four between the .05 and .10 level. With 31 comparisons, less than three would be expected to be significant at the .05 level or beyond by chance alone.

When we relate these differences to changes in the broad areas of functioning with which the study was primarily concerned, i.e., interpersonal relations, utilization of actors' and other group members' resources in problem solution, and effectiveness of the group or individual in achieving specified goals, the results might be ordered as follows:

Nature of interpersonal relations. Groups under threat showed a greater concern with group acceptance and showed less autocratic and aggressive behavior. Other differences for variables included in this class, although nonsignificant, are consistent with these results. Threat groups were more sociable, more cooperative, showed less antagonism, were less competitive, less dominant, showed less aggression, showed a greater degree of informal friendliness, and were characterized by less conflict. A difference not consistent with the above is the greater amount of solidary behavior shown by control groups. Thus, for the 12 variables that reflected differences in the nature of interpersonal relations, 11 are consistent, although only three of the differences were significant.

It might be inferred that these differences indicate that threat results in attempts to maintain interpersonal tensions at a low level. These attempts are characterized by both a reduction in aggressive, antagonistic, competitive behaviors, and an increased striving for group acceptance.

Utilization of resources in problem solution. Groups under threat were less achievement motivated, showed less leadership behavior, gave fewer suggestions, and were less active than control groups. Threat groups asked for orientation from each other more than control groups. As before, other variables reflecting differences in this area of functioning, although nonsignificant, seem consistent with the above. Control groups were more goal oriented, showed a greater degree of equal participation in task solution, gave more suggestions and opinions, and asked less for other group members' opinions.

These differences would appear to indicate

TABLE 7

SIGNIFICANCE OF DIFFERENCES BETWEEN THREAT AND CONTROL CONDITIONS IN SUBJECTS' RESPONSE TO AN ADJECTIVE CHECK LIST*

INDEX	THREAT (N = 36)		CONTROL (N = 9)		F†	t†
	M	σ^2	M	σ^2		
I Anxiety	6.97	15.11	7.11	12.86	1.18	<1
IIa Elated	8.00	14.40	9.33	25.00	1.74	<1
IIb Depressed	1.42	5.62	.67	1.75	3.21	1.27
III Hostile aggression	3.17	9.91	4.00	7.75	1.28	<1
IV Organic symptoms	2.11	4.04	1.00	1.00	4.04‡	2.36‡
V Relaxation	8.95	10.91	9.78	20.44	1.87	<1
VI Positive aggression	21.11	40.96	23.00	35.75	1.15	<1
VIIa Cooperative	18.72	29.06	19.44	36.78	1.27	<1
VIIb Obstructive	6.75	10.31	6.56	10.53	1.02	<1
VIIIa Dominant	13.25	27.85	14.00	20.25	1.38	<1
VIIIb Submissive	5.47	10.31	5.00	2.75	3.75	<1
IX Task involved	20.92	62.08	21.44	52.28	1.19	<1

* There were 114 adjectives, which defined nine indices. An index score was obtained by summing the number of adjectives checked for the particular index. Index scores for the three group members were summed to give a group score.

† Positive values of F and t indicate threat > control; negative values indicate control > threat.

‡ Significant at .05 level or beyond.

TABLE 8

SIGNIFICANCE OF DIFFERENCES BETWEEN THREAT AND CONTROL CONDITIONS IN SUBJECTS' RATINGS OF THEIR GROUPS*

CHARACTERISTIC RATED	THREAT (N = 36)		CONTROL (N = 9)		F†	t†
	M	σ^2	M	σ^2		
1. Goal orientedness	75.03	19.74	74.44	17.03	1.16	<1
2. Productivity	10.53	1.68	10.33	1.00	1.68	<1
3. Degree of equal participation	34.44	23.63	35.11	25.61	−1.08	<1
4. Degree of personality conflict	23.06	23.37	24.00	31.75	−1.36	<1
5. Informal friendliness	52.03	19.91	54.89	10.36	1.92	2.20‡
6. Communication effectiveness	47.36	29.61	49.11	17.86	1.66	1.05
7. Competition between members	26.67	22.97	25.00	12.75	1.80	1.16
8. Cohesiveness	63.14	30.47	63.33	30.00	1.00	<1
9. Satisfaction with session	46.53	18.66	49.33	30.25	−1.62	1.42

* Index scores similar to those obtained from observer ratings were obtained (see Table 6) and summed over the three group members to give a group index score.

† Positive values of F and t indicate threat > control; negative values indicate control > threat.

‡ Significant at .05 level or beyond.

that threat groups are less task oriented and are less forceful, initiating, and active in their attack on the problem than control groups.

Effectiveness of group performance. Groups under threat were less efficient and adaptable than control groups. Supportive, though nonsignificant, is the greater productivity and communication effectiveness of control groups.

An inspection of the tables indicates another finding of considerable interest. Threat groups showed significantly greater variability than control groups in efficiency, productivity, communication effectiveness, adaptability, leadership behaviors, and striving for group acceptance. Although F tests for other variables yielded nonsignificant differences in variance, once again there was a consistency that appears important to consider. For 24 of the 31 variables, threat groups showed greater variability than control groups. Of the seven variables on which control groups showed greater variability, five are concerned with interpersonal relations. It would appear that threat results in considerable variability in task-oriented behaviors and large variability in group effectiveness, but also in consistency of behavior in the interpersonal sphere.

The data presented above are concerned with overt behavioral consequences of exposure to threat. Data concerned with the phenomenal state of *S*s and their reactions to their groups are presented in Tables 7 and 8. Of the 12 comparisons of adjective check list responses, only one was significant at the .05 level or beyond. Threat groups checked more adjectives describing organic symptoms than control groups. Although nonsignificant, it is interesting to note that threat groups describe themselves as less elated, more depressed, and less relaxed than control groups. These differences would tend to confirm our unsystematic observations that threat, as operationally defined, did affect the emotional and/or motivational state of *S*s.

Though nonsignificant, differences for other variables on the adjective check list appear congruent with the differences noted in overt behavior. Threat groups described themselves as less hostile, less positively aggressive, less dominant, and less task involved, as well as more submissive. Since the adjectives were checked *after* the session, these differences might simply indicate that *S*s use the same behavioral cues as *O*s in evaluating their motivational state; or, they might be considered as reflecting emotional concomitants of the behavior changes. Until more is known concerning the determinants of *S*s' responses to various self-description questionnaires, there is little

basis for selecting between the alternative hypotheses.

Of the nine indices concerned with group characteristics only one achieved significance at the .05 level or beyond. Control *S*s rated their groups as being more friendly than threat *S*s. Other differences were small and inconsistent. It should be noted that this instrument appears inadequate because of the extremely small variance obtained as a result of the restricted range of the response scale used by *S*s.

DISCUSSION

The failure to obtain significant results for locus and target can be interpreted in several ways. It might be a reflection of a "real" lack of effect of such variables but, in view of the apparent significance of such factors in nonlaboratory situations, it is more likely a function of the design.[7] The necessity for imposing an external threat in order to obtain *S*s might have precluded effective manipulation of locus and target by a simple variation in instructions at the time of evaluation. It is also possible that the true difference to be expected is small, and a more sensitive design, one in which intergroup variation is reduced, would have resulted in significant differences.

The results obtained for threat versus control conditions might be broadly summarized as follows:

1. Threat appears to result in a reduction of forceful, assertive, aggressive, interpersonal, as well as task-directed behavior, and in a greater concern with group acceptance. These changes might be considered to reflect a tendency to keep tensions within the group at a low level even at the expense of effective task activity.

These results are similar to those obtained by Lanzetta (7), who used a situationally produced stress of a quite different nature than the one utilized in this study, and two dissimilar tasks. Analogous results were also obtained by Wright (20) in a study of children's social relations under frustration.

2. Threat results in poorer group effectiveness and greater variability in effectiveness. This result is similar to those obtained in studies of individual performance under stress. In their recent review of the literature, Lazarus, Deese, and Osler (8) note that the majority of studies have shown impairment of verbal and perceptual motor performance under stress, and several have found greater variability between subjects as a major effect.

Attempts to fit these results into the framework of a contemporary theory are probably premature at this time, but some tentative comments can be offered concerning the possible direction such an assimilation might take.

It might be postulated, first, that increase in anxiety is the primary concomitant of threat, and that further, anxiety will accentuate security operations in an attempt to terminate or diminish the anxiety. The latter postulate has frequently been stated by learning and personality theorists (12, 14). The nature of the security operations is a function of many aspects of the total stimulus situation other than the cues which, themselves, evoke the anxiety. We would postulate that one of the most important aspects is the extent to which potential support is present; more specifically, support of other individuals or a group. If such potential support is perceived, then it seems plausible that an individual will behave in a fashion that will be most likely to assure this support. He will inhibit behaviors that are likely to alienate him from, or cause isolation by, other group members. Thus, we can expect an inhibition of aggressive, self-oriented behavior, which has probably been negatively reinforced, and an increased striving for acceptance.

The assumption of anxiety as a drive that will primarily elicit responses that tend to reduce the drive can also be used to explain the obtained decrease in effectiveness of the group. In the particular task utilized, rapid decisions by the group were required for effective action. When group consensus is required, rapid decision is sometimes best achieved by forceful, initiating behavior by individual group members. Such responses would be incompatible with the responses evoked by the anxiety drive, and thus, if the anxiety drive is strong, would be inhibited. Since ineffective task solution would not reduce *the drive* for task solution, a complete inhibition of task responses would not be expected. Rather, there would occur alternation of such responses with the anxiety-produced responses, owing to an oscillation in

[7] Two more sensitive designs, although for our purposes they appeared experimentally unfeasible, would be a repeat measure or matched group design since they would tend to reduce the variance attributable to group differences.

the relative strength of the anxiety and task-oriented drives. Since the anxiety drive is assumed to be relatively weak under nonthreat conditions, response interference would not be expected; thus, effectiveness should be greater.

The greater variability in effectiveness under threat may be due to the differential strength of the evoked anxiety drive; or, more probably, it may reflect variation in the kind of responses that have been reinforced by anxiety reduction. Rosenzweig's (16) distinction between need-persistive and ego-defensive reactions to frustration comes to mind here. If need-persistive (goal-oriented) responses are high in the habit hierarchy, rather than ego-defensive (protective) reactions, increased anxiety might actually lead to more effective performance.

Recently, Waterhouse and Child have postulated an essentially similar formulation with respect to frustration. Their proposition is: "Frustration will produce a decrease in the quality of on-going performance, to the extent that the frustration evokes other responses which interfere with that on-going performance" (19, p. 298). Substitution of the concept "anxiety" for frustration and additional postulates about the specific nature of the competing responses would allow a posteriori deduction of the obtained results from their proposition. Since they have recently presented experimental support for this proposition, it might tentatively be assumed to have generality for all anxiety-evoking situations.

SUMMARY AND CONCLUSIONS

Thirty-six threat and nine control (nonthreat) groups of three members each worked in a laboratory setting on a task called "intercept." Threat was imposed by structuring the sessions as evaluation tests, the results of which would become part of *S*s' ROTC records. The 36 threat groups were further randomly divided into four classes of nine groups each, each class being assigned to a different cell of a 2 × 2 factorial. The variables of main concern were locus of threat (whether external or internal to the group) and target of threat (whether threat was directed to the group as a whole or to the individual group member).

Records of the ongoing behavior of group members, trait ratings of each member, ratings of group characteristics, and *S*s' perception of their group and their own motivational state were obtained. Analysis of the results was primarily by analysis of variance, and Student's *t* techniques.

Locus and target did not prove to be significant variables, very likely because of the insensitivity of the design. All threat groups were thus combined for comparison with control groups. Threat affected behavior in three main areas of functioning.

1. Interpersonal relations: Threat groups showed less behavior that would result in tension and friction within the group and more behavior oriented toward group acceptance.

2. Utilization of actors' and other group members' resources: Threat groups were less task oriented and were less forceful, assertive, and active in their attack on the task problem. They also showed greater variability in such behavior.

3. Effectiveness of the group: Threat groups were both less efficient and adaptable and more variable in effectiveness.

An attempt was made to interpret these results within the framework of a competing response theory of anxiety.

REFERENCES

1. BALES, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley, 1950.
2. CARTER, L., HAYTHORN, W., SHRIVER, BEATRICE M., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589–595.
3. CARTER, L., HAYTHORN, W., MEIROWITZ, BEATRICE, & LANZETTA, J. The relation of categorizations and ratings in the observation of group behavior. *Hum. Relat.*, 1951, **4**, 239–254.
4. EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
5. ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57**, No. 3 (Whole No. 263).
6. KOGAN, L. S., & HUNT, J. MCV. Problems of multijudge reliability. *J. clin. Psychol.*, 1950, **6**, 16–19.
7. LANZETTA, J. Group behavior under stress. *Hum. Relat.*, in press.
8. LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.
9. LINDSLEY, D. B., & OTHERS. A study of performance under speed stress. (OSRD, 1945; Publ. Bd., No. 18381.) Washington, D. C.: U. S. Dep. Commerce, 1946.
10. MCCLELLAND, D. C., & APICELLA, F. S. A functional classification of verbal reactions to experimentally induced failure. *J. abnorm. soc. Psychol.*, 1945, **40**, 376–390.

11. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
12. MAY, R. *The meaning of anxiety*. New York: Ronald, 1950.
13. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435–472.
14. MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
15. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
16. ROSENZWEIG, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944. Pp. 379–388.
17. SHRIVER, E. L., & SHRIVER, BEATRICE M. A demonstration of a modified analysis of variance used to obtain comparable reliability measures of observers, groups, and individual subjects in group dynamics data. Paper read at Eastern Psychol. Ass., Boston, April 1953.
18. TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
19. WATERHOUSE, I. K., & CHILD, I. L. Frustration and the quality of performance: III. An experimental study. *J. Pers.*, 1953, **21**, 298–311.
20. WRIGHT, M. E. The influence of frustration upon the social relations of young children. *Charact. & Pers.*, 1943, **12**, 111–122.

Received April 3, 1954. Prior publication

A STUDY IN LANGUAGE AND COGNITION[1]

ROGER W. BROWN

Harvard University

AND ERIC H. LENNEBERG

Center for International Studies, Massachusetts Institute of Technology

IT IS popularly believed that reality is present in much the same form to all men of sound mind. There are objects like a house or a cat and qualities like red or wet and events like eating or singing and relationships like near to or between. Languages are itemized inventories of this reality. They differ, of course, in the sounds they employ, but the inventory is always the same. The esthetic predilections of the Italian lead him to prefer euphonious vowels, while the German is addicted to harsh consonant groupings, but the things and events named are the same in both tongues. We are confirmed in this view by our linguistic education, which requires us to memorize lists of French or German or Latin words and their exact English equivalents.

There are, of course, poetic persons who claim to find in each language some special genius that peculiarly fits it for the expression of certain ideas. But the majority of us are at a loss to understand how this can be, since there is apparently a relationship of mutual translatability among the languages we learn. To be sure, we can see that one language might contain a few items more than another. If the Germans were to invent a new kind of automobile and we had not yet thought of such a machine, their dictionary would have one entry more than ours until we heard of the discovery and named it for ourselves. But these inequalities are in the lexical fringe. They do not disturb the great core of common inventory.

THE WHORF THESIS

This linguistic ethnocentrism will be seriously disturbed by the study of languages that lie outside the Indo-European group. It has not prepared us for finding that there is a language in which noun and verb categories apparently do not exist, or that there is another in which the colors gray and brown are called by the same name. Such data from the study of American Indian tongues led Whorf (18) to reject the usual view of the relationship between language and thought. He suggested that each language embodies and perpetuates a particular world view. The speakers of a language are partners to an agreement to see and think of the world in a certain way—not the only possible way. The world can be structured in many ways, and the language we learn as children directs the formation of our particular structure. Language is not a cloak following the contours of thought. Languages are molds into which infant minds are poured. Whorf thus departs from the common sense view in (*a*) holding that the world is differently experienced and conceived in different linguistic communities and (*b*) suggesting that language is causally related to these psychological differences.[2]

Other authors have believed that the relationship between language and thought is somewhat as proposed by Whorf. Cassirer (2), the distinguished philosopher, maintained that language is the direct manifestation of knowledge; he explicitly denied a form-content relationship between words or language structure and isolates of knowledge. In this he was in agreement with such other German writers as Wundt (19) and Bühler (1). Orwell (15) in his novel *Nineteen Eighty-Four* describes a totalitarian England of the future. The really efficient dictatorship of that day invents a language—Newspeak—in which it is impossible not only to express, but even to think, a rebellious thought. An equally great faith in the causal efficacy of language lies behind the General Semantics movement. Korzybski (8), for instance, holds that clear thinking and so-

[1] This project was carried out under the joint auspices of the Communications Program of the Center for International Studies, Massachusetts Institute of Technology, and the Laboratory of Social Relations, Harvard University.

[2] While this seems a fair statement of Whorf's usual views, he occasionally took a somewhat more conservative position.

cial progress are to be attained through the reform of language.

Cognitive Differences between Linguistic Communities

The first tenet of the Whorf thesis is that the world is differently experienced and conceived in different linguistic communities. The evidence presented in support of this claim is entirely linguistic. It will be helpful to distinguish between the conclusions based on lexical features of two languages and those based on structural features.

Lexical features. In the Eskimo lexicon there are three words to distinguish three varieties of snow. There are no single-word equivalents for these in English. The word "snow" would be used to describe all three. What psychological conclusions can be drawn from these data? Does the Eskimo see differences and similarities that we are unable to see?

Psychologists ordinarily infer perceptual discrimination when a subject is consistently able to respond differently to distinctive stimulus situations. The subject may be rat, dog, or man. The response may be running, salivation, or—speech. Words are used meaningfully when they are selectively employed with reference to some kind of environment—whether physical, social, or linguistic. The linguist in the field may discover the referent of a term by noting the pattern of its usage. The Eskimo's three "snows" are sufficient evidence from which to infer that he discriminates three varieties of snow. These selective verbal responses satisfy the conditions for inferring perceptual discrimination.

What can be said of the English speaker's ability to distinguish the same three kinds of snow? When different stimuli do not elicit differential responses, the stimuli may or may not be discriminated. A subject may be perfectly able to distinguish two situations and still not care to do anything about it. Consequently the fact that English speakers do not have different names for several kinds of snow cannot be taken to mean that they are *unable* to see the differences. It would seem, then, that all such comparisons are psychologically inconclusive. The Eskimo and American may or may not see the world differently.

There is, however, other evidence to indicate that the speaker of English can classify snows as the Eskimo does. If we listen to the talk of small boys, it is clear that they perceive at least two kinds of snow—good-packing snow and bad-packing snow. This is a distinction of the greatest importance to anyone interested in making snowballs. This discrimination is evidenced by differential response—not distinct lexical items but combinations of items—"good-packing snow" and "bad-packing snow." Whorf himself must have been able to see snow as the Eskimos did since his article describes and pictures the referents for the words. Since both Eskimo and American are able to make differential responses to snows, we must conclude that both are able to see differences. This seems to lead us to the conclusion that the Eskimo and American world views do not differ in this regard.

Although the three kinds of snow are namable in both Eskimo and English, each of them requires a phrase in ordinary English, whereas single words will do it for the Eskimo. Zipf (20) has shown that there exists a tendency in Peiping Chinese, Plautine Latin, and American and British English for the length of a word to be negatively correlated with its frequency of usage. This is true whether word length is measured in phonemes or syllables. It is not difficult to find examples of this relationship in English. New inventions are usually given long names of Greek or Latin derivation, but as the products become widely known and frequently used in conversation the linguistic community finds shorter tags for them. Thus the "automobile" becomes the "car" and "television" shrinks first to "video" and eventually to "TV." Three-dimensional movies are predictably described as "3-D."

Doob (3) has suggested that this principle bears on Whorf's thesis. Suppose we generalize the findings even beyond Zipf's formulation and propose that the length of a verbal expression provides an index of its frequency in speech and that this, in turn, is an index of the frequency with which the relevant perceptual judgments of difference and equivalence are made. If this is true, it would follow that the Eskimo distinguishes his three kinds of snow more often than Americans do. It would mean—to cite another example—that the Hopi is less often called upon to distinguish airplanes, aviators, and butterflies than is the American, since the Hopi has but a single name for all

three of these. Such conclusions are, of course, supported by extralinguistic cultural analysis, which reveals the importance of snow in the Eskimo's life and the comparative indifference of the Hopi to airplanes and aviators.

We will go further and propose that increased frequency of a perceptual categorization will mean a generally greater "availability" of that category. In the experimental study of memory we are accustomed to think of the methods of recall, recognition, and relearning as increasingly sensitive indices of retention. In the experimental study of categorizing behavior there are two principal methods: (*a*) Goldstein's (6) technique of presenting a subject with an array of objects and asking him to group them, and (*b*) Hull's (7) discrimination learning technique. Hull's method seems to be the more sensitive of the two. We should guess that when the Eskimo steps from his igloo in the morning and is confronted by a snowy world, these snows will fall into named categories for him in a way that they will not for the American. If, however, the American were subjected to a discrimination learning experiment, or if the perceptual structure were otherwise made worth his while, he could see snow as does the Eskimo. We think, really, that more namable categories are nearer the top of the cognitive "deck."

Structural features. Members of structural categories have no phonetic common denominator. They are grouped together because they have the same structural relations with other forms in the language. In English, nouns constitute a structural category; its members can appear with definite and indefinite articles, can form plurals in certain ways, etc. In French all nouns of the feminine gender belong to one structural category since they all require the feminine articles and suffixes.

Whorf generally assumes that structural categories are also symbolic categories. When he finds structural differences in languages he concludes that there are parallel cognitive differences. There are in Hopi two structural categories showing some similarity to our verb and noun categories, with the difference that one of the Hopi classes includes only the names for such short-term events as lightning, flame, and spasm, while the other includes only such long-term events as man, house, and lifetime. Whorf concludes that the Hopi organizes his world on a dimension we usually overlook. When the structural class has such obvious semantic properties, Whorf's conclusions have a kind of plausibility.

However, very few structural classes have such clear and consistent meanings. In the languages we know best, those of the Indo-European family, there are many structural categories with no discernible meaning. In French, for instance, it is not clear that the gender of a form signifies anything to a speaker. Certainly it is difficult to find any common attributes in the references for French nouns of feminine gender. Not even the majority of them manifest feminine sexuality—even in the extended sense of Freud. The French speak of "*le* balcon" in spite of their saying, "Elle a du balcon." The linguist Charles Fries (5) has shown how difficult it is to describe a semantic for the English "parts of speech." If the noun can be defined as "the name of a person, place, or thing," this is only because "thing" is left unexplicated. It serves handily to designate whatever is nominalized and yet neither person nor place.

Even where the ethnolinguist can discover consistent structural meanings, it does not follow that these meanings are present to the native speakers of a language. Suppose that a subject in the laboratory were required to signal his recognition of each of ten different musical chords by raising that one of his ten fingers which has been designated for each chord. If all extraneous sensory information were excluded, his ability to pattern correctly the movements of his fingers would be evidence of his ability to identify the chords. The experimenter might introduce a potential structural meaning by ruling that the fingers of the right hand would always be raised for chords in the major mode and the fingers of the left hand for minor chords. The subject's responses might follow this pattern and yet he need never have detected the major and minor modes. Similarly, even if there were some semantic to French gender, one could speak the language without detecting it. "La fille" and "la femme" could be learned without noticing that both are in the feminine mode. No safe inferences about cognition can be made on the basis of the simple existence of the structural classes described by Whorf. The structural evidence is extremely difficult to interpret, and it seems

clear that psychological data are needed to supplement those of the linguist.

Language in Causal Relation to Cognition

The second major tenet of Whorf's thesis is that language causes a particular cognitive structure. In what way can this occur? There seem to be two possibilities. Suppose that the colors red and green are not "given" categories but must be learned. A father who has formed these categories may play a game with his child that will teach the categories. The green blocks are to be used for building a house and the red ones for a barn. The child cannot properly pattern the blocks without learning to make the visual distinctions involved. Notice that the barn and house are not essential here. A father could ask his child to tell him whether each block is red or green. In learning this game, too, the child necessarily would learn to perceive the colors. Because words have symbolic properties, because their usage is patterned with reference to the total environment, language can cause a cognitive structure. To the degree that children are motivated to speak a language as it is spoken in their community they are motivated to share the world view of that community. To be sure, linguistic training is not the only means of procuring cognitive socialization; the house-barn game demonstrates that. The word game has the tremendous advantage that it can be played constantly and concurrently with many other activities. The child and his adult tutor can chatter together whether they are walking or riding, playing or working. In this chatter more is taught than a simple motor skill involving the muscles of articulation. A total culture is internalized.

There is a second, more dubious, avenue for the influence of language on thought. If life is a river, speech is a babbling brook whose course parallels that of the river. The brook is smaller and simpler than the river. A child can learn the phonemic structure of his language fairly easily. He will also realize that as the phonemic patterns he hears spoken change there are important changes in the nonlinguistic world. There is, for instance, an important difference that goes with the shift of speech from "father" to "mother." When, on the other hand, combinations of phonemes are repeated, two situations are equivalent in some important way. Consider the "strike" and the "ball" in baseball. These are rather difficult categories. The differences between them are subtle and complex. A naive observer of a baseball game would have a difficult time learning these categories by simply observing the game. It makes a great difference that the umpire calls out "strike!" each time a member of that category occurs and "ball!" to identify an instance of the other category. The umpire's shout directs us to look here and now to discover something of importance. The word spotlights a moment of consciousness and puts it in connection with other events similarly spotlighted. The various "strikes" are equivalent in some way and distinct as a category from the events labelled "ball." The babbling brook can, then, be a guide to the structure of the more complex but also more interesting river.

All of our reasoning cannot be said to prove the validity of any set of psychological conclusions. It does, however, point the direction for such a proof and suggests empirical steps that will advance our knowledge of this problem. We have made a small beginning in this work.

Our findings bear on only one of the claims made by Whorf—that there are cognitive differences correlated with lexical differences. We have developed lexical differences into the variable of "codability" and attempted to spell out the relationship between this variable and a single cognitive performance—recognition.

The Experiment

Sensory psychologists have described the world of color with a solid using three psychological dimensions: hue, brightness, and saturation. The color solid is divisible into millions of just noticeable differences; *Science of Color* (14) estimates 7,500,000. The largest collection (4, 11) of English color names runs to less than 4,000 entries, and of these only about 8 occur very commonly (17). Evidently there is considerable categorization of colors. It seems likely to us that all human beings with normal vision will be able to make approximately the same set of discriminations. This ability appears to depend on the visual system, which is standard equipment for the species. Whatever individual differences do exist are probably not related to culture, linguistic or extralinguistic. It does not follow that people everywhere

either see or think of the color world in the same way. Cultural differences probably operate on the level of categorization rather than controlled laboratory discrimination.

Our explorations in the Yale Cross-Cultural Index turned up many reports of differences on this level. Seroshevskii (16), for instance, has reported that in the Iakuti language there is a single word for both green and blue. This is the kind of language difference discussed in the first section of this paper. A region of experience is lexically differentiated in one culture but undifferentiated in another. Color categories differ from such categories as snows in that they have boundaries that can be plotted on known dimensions. Color categories, furthermore, are continuous with one another, sharing their boundaries. Consider for a moment the single dimension of hue taken at a high level of saturation and brightness. Native speakers of English could be shown various shades and asked to give the usual color name for each stimulus presented. For each common color name there would be some shades invariably called by that name. There would be other shades sometimes associated with one name, sometimes with another. When the responses are divided about equally between two or more names, we should have boundaries between categories. If a native speaker of Iakuti were asked to provide the usual color names for the various shades, we should anticipate a somewhat different pattern. English speakers would have trouble naming the hues in the boundary region between green and blue. Probably they would hesitate, disagree among themselves, and sometimes use phrases or such combination names as "greenish blue." For the Iakuti, on the other hand, this region is right in the center of a category and would be named with great ease.

Of course, our example is greatly simplified over the actual case since we have dealt with the single dimension of hue whereas the color lexicon is actually patterned with respect to all of the three dimensions of visual experience. When these are considered, the range of applicability of a color term is a space within the color solid rather than a distance along a line. The simplification was for expository purposes and does not alter the logic of the argument.

This example of a cultural difference serves to introduce the variable *codability*. Certain colors are differentially codable in the Iakuti and English languages. So long as the data collected are of the usual linguistic variety, this difference of codability will be manifest in only one way—environmental distinctions expressed lexically in one language are expressed with word combinations in another language. Our reasoning led us to expect differential availability of reference categories in such a case. We undertook experimental work to discover additional behavioral indices of codability, and hoped to find one more sensitive than that which can be teased out of linguistic data. If we found such an index, we would go on to explore the behavioral consequences of differential availability of cognitive categories.

There are differences of codability within English itself. Some shades fall safely within the province of a given name while others lie within boundary regions. Here it is a matter of comparing the English codability of one region of visual experience with another region, whereas the ethnolinguist has usually compared the codability of one region of experience in several languages. If we explore the codability variable in English, it seems likely that our discoveries will apply to the cultural differences with which the inquiry began. If a general law can be found relating codability to availability, individual cultures may conform to the law though they differ among themselves in the values these variables assume in particular regions of experience.

Measurement of Codability

The entire series of Munsell colors for the highest level of saturation ("chroma" as Munsell calls it) was mounted on cards in systematic fashion. Five judges were asked to pick out the best red, orange, yellow, green, blue, purple, pink, and brown from these 240 colors. These names are the most frequently appearing color terms in English (17). For each name the color chip most often selected was added to our test list. Agreement among judges was high, and it is quite clear, therefore, that there is in this series one particular color chip with the best claim to each color name. The number of colors was then raised to 24 by adding chips that would, in combination with the first 8, provide as even a coverage of the color space as practicable. These colors are specified in Table 1. One set of the 24 chips was mounted on white 3 × 5 cards, one chip to a card. Another set was arranged randomly on a single large card.

To expose the single small cards a drop shutter was mounted in a 3 × 2-foot gray (Munsell neutral value 6, reflectance 30 per cent) board. The board was about three feet from the subject's (*S*'s) eyes and was

TABLE 1

THE MUNSELL NOTATION AND SCORES FOR DISCRIMINABILITY, CODABILITY, AND RECOGNITION FOR THE 24 TEST COLORS

MUNSELL NOTATION*	DISCRIMINABILITY		CODABILITY		RECOGNITION (GROUP C TABLE 3)	
	SCORE	RANK	SCORE	RANK	SCORE	RANK
2.5R 7/8	38	2	18	9.5	.875	8
2.5R 5/10	27.5	6	7	18.5	.694	11
5R 4/14	23	10.5	19	7.5	1.020	5
7.5R 8/4	18	15	7	18.5	.236	18
2.5YR 6/14	38	2	29	1.5	1.499	2
5YR 3/4	24	9	26	3	.972	7
7.5YR 5/8	26	7.5	8	16	.736	9
2.5Y 7/10	12	19	3	24	.486	13
5Y 8/12	37	4	25	4	2.450	1
7.5Y 6/8	13	17	4	23	.250	17
3GY 7.5/11.2	23	10.5	14	12	1.222	4
7.5GY 3/4	9.5	23	14	12	0.000	23.5
2.5G 5/8	18.5	14	23	6	.986	6
7.5G 8/4	17.5	16	19	7.5	.167	19
5BG 3/6	4.5	24	12	15	.111	22
10BG 6/6	21	12	7	18.5	.458	14
8.5B 3/6.8	38	2	13	14	0.000	23.5
2.5PB 7/6	19	13	18	9.5	.436	16
5PB 4/10	10.5	21	29	1.5	.695	10
10PB 5/10	12	19	7	18.5	.125	20.5
5P 8/4	12	19	14	12	.547	12
10P 3/10	10	22	24	5	.444	15
5RP 6/10	26	7.5	6	21.5	.125	20.5
8RP 3.4/12.1	31	5	6	21.5	1.464	3

* For conversion to C.I.E. Tristimulus values and Source C, C.I.E. chromaticity coordinates see Nickerson, Tomaszewski, and Boyd (13).

illuminated from above and behind by a General Electric standard daylight fluorescent lamp.

The *Ss* were 24 Harvard and Radcliffe students who spoke English as a native language and had no particular training in distinguishing colors. They were screened for color blindness with the standard Pseudo-Isochromatic Plates.

The *Ss* were first shown the 24-color random chart for about five minutes. After the chart was removed, they were told that each of the colors on the chart would appear individually in the tachistoscope and that *S*'s task was to give the name of each as it appeared. "Name" was defined as the word or words one would ordinarily use to describe the color to a friend. The *Ss* were urged to be both quick and accurate.

The 24 colors were presented in a predetermined random order for each *S*. No order was repeated. Each color was exposed until *S* had named it. In our trial procedure we used a voice key and chronoscope to measure the reaction time. The scope was activated by the opening shutter of the tachistoscope and stopped by *S*'s first vocalization. This method proved to be unsuitable since *Ss* would frequently burst out with something other than a color name, which, of course, stopped the undiscriminating chronoscope. Consequently, we abandoned this technique and used the stop watch. The watch was started as the experimenter (*E*) dropped the shutter and stopped at first mention of a color name.

The variable of codability was measured in five ways. (*a*) The average length of naming response to each color was obtained by counting syllables. (*b*) The average length was also obtained by counting words. (*c*) The average reaction time for each color was obtained by ranking all of the reaction times of an individual *S* and taking the mean rank across *Ss* for every color. (*d*) The degree to which *Ss* agreed with one another in naming a color was assessed as follows: We counted the total number of different responses to a color (DR) and also the number of *Ss* who agreed on whatever response was most often given to a particular color (CR). The first value was subtracted from the second and a constant of 20 added to keep the results positive (CR − DR + 20). Color 18, for example, was given the following eight different names: gray-blue, blue, light gray-blue, light blue, very pale blue, light blue-gray, pale blue, and powder blue. Of these, the single-word response "blue" occurred most often—six times. Color 18, then, scored 6 − 8 + 20, or 18. (*e*) The degree to which *Ss* agreed with themselves from one time to another in naming a color was calculated as follows: Five *Ss* were recalled after a period of one month and subjected to a repetition of the naming procedure. When an *S* gave identical responses to a color on the two occasions, we counted one agreement. We determined the number of agreements for each *S* and considered that to be unity. Each individual agreement was then given the appropriate fractional value. Suppose an *S* had eight agreements. If he agreed in his name for Color 11, he would add $\frac{1}{8}$ to the score for that color. The agreement score is, then, the sum of the individual performances weighted for each individual's over-all tendency to agreement.

In Table 2 the intercorrelations of scores on these five measures appear. All correlations are in the predicted direction and most of them are significant, with .355 the smallest. With a single iteration this matrix yielded a general factor which we call codability. No correlations over .113 remain after the extraction of this single factor. Our fourth index, the degree of agreement between *Ss*, has by far the largest factor loading. It was selected as the measure of codability for the second phase of the

TABLE 2

CORRELATION MATRIX FOR FIVE INDICES OF CODABILITY

MEASURE	1	2	3	4	5
Number of syllables					
Number of words	.425*				
Reaction time	.387	.368			
Interpersonal agreement	.630*	.486*	.864*		
Intrapersonal agreement	.355	.537*	.649*	.773*	
k from second factoring	.589	.587	.787	.976	.795
Communality from first factoring	.403	.378	.671	.873	.653

* $p \leq .05$.

experiment. The obtained codability values for the 24 colors are listed in Table 1.

Codability and Recognition

Once the codability variable suggested by Whorf's ethnolinguistic observations had been operationalized, it remained to relate this variable to some nonlinguistic behavior which might be considered an index of availability. We selected the recognition of colors.

From the 240 Munsell chips taken at highest saturation we selected out alternate chips, taking care to include the 24 colors for which codability data had been collected. The resultant collection of 120 colors was systematically mounted on a white card. Hue varied along the vertical dimension of the card and brightness on the horizontal dimension. Since there were 20 steps of hue and only 6 of brightness, we divided the total colors in half and mounted one half above the other so as to make a more manipulable display.

New *S*s were screened, as before, for color blindness and language background. The basic procedure was to expose simultaneously 4 of the 24 colors, remove them, and ask *S*s to point to the colors just seen, on the large chart of 120. Neither *E* nor *S* mentioned any color name during the session. The recognition score for a color was computed as follows: We determined the number of correct identifications made by each *S* and considered this number to be unity. Each individual correct identification was given the appropriate fractional value. Suppose for instance, that an *S* who correctly identified a total of six colors recognized Color 24. This recognition would have counted as ⅙ on the total recognition score for that color. Another *S* for whom Color 24 was one of eight correctly identified colors would have contributed ⅛ to the score for Color 24. In other words, the recognition score for a color is the sum of the individual performances weighted for each *S*'s over-all ability to recognize colors. The scores for the 24 colors appear in Table 1.

In trial runs, *S*s were asked how they managed to retain the four colors in memory after they were removed from sight. Most *S*s reported that they named the colors when they were exposed and "stored" the names. It seemed likely, therefore, that those colors that are quite unequivocally coded would be most likely to be recognized. When a color elicits a considerable range of names, the chances of recovering the color from the name would be reduced. This expectation was fulfilled by a rank-order correlation of .415 between codability and recognition scores.

There is, however, another variable that influenced recognition. The 120 colors used are not perceptually perfectly equidistant. The manufacture of equidistant color chips is technically difficult and expensive and, indeed, above a certain level of saturation, impossible. Since we were unable to control experimentally the variable "discriminability," we must ask whether or not our findings were due to a positive correlation between codability and discriminability. Could it be that our codable colors were so distant, perceptually, from their nearest neighbors that their superior recognizability was actually due to these better discrimination conditions? To obtain an answer to this question we determined the true perceptual distance between each of the colors used from the Newhall, Nickerson, and Judd (12) charts. These charts convert every Munsell book notation into a renotation which is the specification of a true perceptual locus of each color within the Munsell coordinate system. The difference between two renotations expresses quantitatively the perceptual distance between the colors.

For each of the 24 test colors we computed a discriminability score which describes its distinctiveness from the colors surrounding it. The difference between two renotations yields three numbers, one for each dimension. To make these numbers perceptually commensurable (i.e., to reduce them to a common denominator), the Optical Society of America Subcommittee on the Spacing of the Munsell Colors suggests the values 3, 2, and 1 for hue, chroma, and value, respectively. Since every color has two neighbors on each of the three dimensions, a total of six numbers will express, in a rough way, the discriminability of that color. The sum of these yields the unadjusted discriminability score. Adjustments of this score are necessary (*a*) because if a color appears on the margin of our chart it has a lower chance of being recognized correctly and (*b*) because a color that has a very close neighbor on one side and distant neighbors on three others might come out with a good discriminability score although the close contiguity on one side would hinder correct recognition considerably. Consequently, colors appearing on the margin of our chart had the constant 3 subtracted from their unadjusted discriminability score, and colors with a close neighbor had the constant 6 subtracted.

Our scoring method is to a certain degree arbitrary, to be sure, but since the equation of perceptual distances on different visual dimensions is an unsolved problem, there seems to be no more objective method available. In addition, of course, all decisions were made without knowledge of recognition scores.

Since we were unable to control discriminability experimentally, we controlled it statistically. The partial correlation between codability and recognition, with discriminability constant, is .438. Furthermore, the correlation between codability and discriminability is .074, which is not significant. Evidently the relation between codability and recognition is not a consequence of variations in discriminability.

Since the reports of our early *S*s indicated that colors were stored in linguistic code, it seemed likely that color codability would increase in importance as the storage factor was maximized in the recognition situation. Discriminability, on the other hand, should remain at the same level of importance or possibly decline somewhat. If, for example, a single color were exposed, removed, and then identified with minimal delay, *S*s

TABLE 3

RECOGNITION PROCEDURES

GROUP	*N*	NUMBER OF COLORS ORIGINALLY EXPOSED	LENGTH OF INTERVAL	CONTENT OF INTERVAL
A	9	1	7 seconds	
B	9	4	7 seconds	
C	16	4	30 seconds	
D	9	4	3 minutes	Tasks

Note.—Exposure time for all groups was 3 seconds.

TABLE 4

Correlations Involving Scores on Codability (C), Discriminability (D), and Recognition (R) with Four Experimental Conditions for Recognition

Group	C with R	D with R	C with R, D constant
A	.248	.540*	.248
B	.411	.460*	.426*
C	.415	.503*	.438*
D	.487*	.505*	.523*

* $p \leq .05$.

might retain some direct memory of the color, perhaps as a visual image. In this situation discriminability would be a determinant of recognition but codability would not be. However, when the number of colors is increased and the interval prolonged and filled with activity, the importance of linguistic coding should increase. Table 3 describes the experimental variations we used. Groups A, B, C, and D are arranged in what we believed to be an order of increasingly difficult storage of colors. Group C is our major group, for which results have already been described. The tasks which filled the interval for Group D were simple but absorbing—the kind of thing often used in experiments on the Zeigarnik phenomenon.

It can be seen from the data in Table 4 that the correlation between recognition and codability scores does increase as the importance of storage in the recognition task increases.[3] The particular order obtained would occur by chance only once in 24 times.

Table 4 also shows that discriminability is most closely related to recognition in Group A, for which the possibility of some direct memory of the color is maximized. The importance of discriminability declines slightly but not significantly as the recognition is made more difficult. Our expectations with regard to both codability and discriminability are generally confirmed.

In the first section of this paper we concluded our discussion of lexical differences between languages with the prediction that a given set of cognitive categories will be more available to the speakers of a language that lexically codes these categories than to the speakers of a language in which the categories are not represented in the lexicon. Lexical differences have been expanded into the variable of codability, and category availability has been operationalized as a recognition score. We found that differences in the English codability of colors are related to differences in the recognition of these colors. We expected these results to apply to the cross-cultural case, and some confirmation of this expectation is available in the results of a study by Lenneberg and Roberts (10). This study of Zuni Indians used a field adaptation of our methods and apparatus. The Zuni color lexicon codes the colors we call orange and yellow with a single term. Monolingual Zuni *Ss* in their recognition task frequently confused the orange and yellow colors in our stimulus set. Our English-speaking *Ss* never made this error. It is a distinction which is highly codable in English and highly uncodable in Zuni. Interestingly, bilingual Zunis who knew English fell between the monolingual Zuni and the native speaker of English in the frequency with which they made these errors.

The Whorf thesis claims more than a simple relationship between language and cognition; language is held to be causally related to cognitive structure. Our correlational evidence does not, of course, establish the direction of causality. If we may be permitted a guess it is that in the history of a culture the peculiar features of the language and thought of a people probably develop together.

In the history of an individual born into a linguistic community the story is quite different. The patterned responses are all about him. They exist before he has the cognitive structure that will enable him to pattern his behavior in the approved fashion. Simple exposure to speech will not shape anyone's mind. To the degree that the unacculturated individual is motivated to learn the language of a community, to the degree that he uses its structure as a guide to reality, language can assume a formative role.

Summary

The Whorf thesis on the relationship between language and thought is found to involve the following two propositions: (*a*) Different linguistic communities perceive and conceive reality in different ways. (*b*) The language spoken in a community helps to shape the cognitive structure of the individuals speaking that language. The evidence for the first proposition derives from a comparison of the lexical and structural characteristics of various languages. The linguistic comparisons alone do not establish the proposition. They need to be complemented with psychological data. The

[3] Kurtz and Hovland (9) have shown that verbalization during observation of stimulus objects facilitates recognition under certain circumstances.

second proposition is not directly supported by any data. However, it is clear that language can be described as a molder of thought since speech is a patterned response that is learned only when the governing cognitive patterns have been grasped. It is also possible that the lexical structure of the speech he hears guides the infant in categorizing his environment. These matters require empirical exploration.

An experiment is described which investigates a part of proposition *a*—the idea that lexical differences are indicative of cognitive differences. Whorf reports many cases in which a given range of experience is lexically differentiated in one language whereas the same discriminations can only be described with phrases in another language. Rather than compare members of different linguistic communities, we chose to work with native speakers of English and to compare their linguistic coding of two regions of experience. Within the realm of color vision there are colors that can be named with a single word and others that require a phrase. This kind of linguistic difference in the length of name (measured by words or syllables) was found to be correlated with the latency of the naming response and the reliability of the response from person to person within the linguistic community and from time to time in one person. A factor analysis of these measures yielded a single general factor—codability. The measure carrying the largest factor loading was the reliability of naming response between individuals who speak the same language. This variable—the codability of a color—proved to be related to *Ss*' ability to recognize colors. Codability accounted for more variance in the recognition task as the task was delayed and complicated to increase the importance of the storage factor. Data obtained from the Zuni Indians show a similar relationship between codability and recognition. It is suggested that there may be general laws relating codability to cognitive processes. All cultures could conform to these laws although they differ among themselves in the values the variables assume in particular regions of experience.

REFERENCES

1. BÜHLER, K. *Sprachtheorie.* Jena: G. Fischer, 1934.
2. CASSIRER, E. *The philosophy of symbolic forms.* Vol. 1. *Language.* New Haven: Yale Univer. Press, 1953.
3. DOOB, L. W. *Social psychology.* New York: Holt, 1952.
4. EVANS, R. M. *An introduction to color.* New York: Wiley, 1948.
5. FRIES, C. C. *The structure of English.* New York: Harcourt, Brace, 1952.
6. GOLDSTEIN, K. *Language and language disturbances.* New York: Grune & Stratton, 1948.
7. HULL, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, **28**, No. 1 (Whole No. 123).
8. KORZYBSKI, A. The role of language in the perceptual processes. In R. R. Blake & G. V. Ramsey (Eds.), *Perception; an approach to personality.* New York: Ronald, 1951. Pp. 170–205.
9. KURTZ, K. H., & HOVLAND, C. I. The effect of verbalization during observation of stimulus objects upon accuracy of recognition and recall. *J. exp. Psychol.*, 1953, **45**, 157–164.
10. LENNEBERG, E. H., & ROBERTS, J. M. The denotata of color terms. Paper read at Linguistic Society of America, Bloomington, Indiana, August, 1953.
11. MAERZ, A., & PAUL, M. R. *A dictionary of color.* New York: McGraw-Hill, 1930.
12. NEWHALL, S. M., NICKERSON, D., & JUDD, D. B. Final report of the OSA subcommittee on the spacing of the Munsell colors. *J. opt. Soc. Amer.*, 1943, **33**, 385–418.
13. NICKERSON, D., TOMASZEWSKI, J. J., & BOYD, T. F. Colorimetric specifications of Munsell repaints. *J. opt. Soc. Amer.*, 1953, **43**, 163–171.
14. OPTICAL SOCIETY OF AMERICA, COMMITTEE ON COLORIMETRY. *The science of color.* New York: Crowell, 1953.
15. ORWELL, G. *Nineteen eighty-four.* New York: Harcourt, Brace, 1949.
16. SEROSHEVSKII, V. R. *Iakuti.* St. Petersburg: Royal Geographical Society, 1896.
17. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.
18. WHORF, B. L. *Four articles on metalinguistics.* Washington: Foreign Service Institute, 1950.
19. WUNDT, W. *Völkerpsychologie.* Vol. 1. *Die Sprache.* Leipzig: Engelmann, 1900.
20. ZIPF, G. K. *The psycho-biology of language.* Boston: Houghton Mifflin, 1935.

Received April 12, 1954. Prior publication.

CRITIQUE AND NOTES

SOME PSYCHODYNAMIC ASPECTS OF FELT PARENTAL ALLIANCE IN YOUNG CHILDREN[1]

EDWARD M. BENNETT AND DOROTHEA E. JOHANNSEN

Tufts College

THE second card of the Children's Apperception Test (1, 2) shows a small bear with a large bear pulling on a rope against another large bear. In a study of 58 diabetic children between the ages of seven and nine, some children used this card to verbalize a conflict, friendly or otherwise, among the members of a family group. Thirty-seven children, 13 boys and 24 girls, specified that the child bear joined either the mother or the father in this conflict. Twenty-four of these 37 children indicated specifically whether the child's side won or lost.

No attempt is made to generalize this projective alliance beyond the possibility that it reflects some form of felt alliance on the part of the child. It is the relationship of this felt alliance to other factors that is presented here, as a research finding needing further, more direct, study for any real generalization.

The findings reported in Table 1 reflect the statistically significant relationships disclosed when the children who allied with the parent of the opposite sex are compared with the children who allied with the parent of the same sex. All tests of significance are based on two by two chi square.

For the 24 children who indicated whether the child's side won or lost the conflict, Finding 1 reflects the presence of a significant tendency for the opposite-sex allied children cognitively to win their symbolic feud, while the same-sex allies show more of a tendency to lose.

Findings 2, 3, 4, and 5 in Table 1 are based on ratings or rankings in which the subject was classified as being *high* or *low* on the variable in question on the basis of his being above or below the median of the total 58 children in the major research sample. During the course of the summer, the children were observed by camp counselors and an in-field psychological assistant. Findings 2 and 3 refer to mean ratings, based on five-point scales, by counselors and the psychological assistant. Finding 4 refers to the mean proportional ranking of the subject, when ranked weekly with the other members of the research sample, by the psychological assistant. Finding 5 is based on the rating of the psychological assistant alone, and is open to experimental error.

Using the general wording of the rating and ranking schedules, Finding 2 indicates that opposite-sex allied children appear more conforming to that which is expected of them. Finding 3 reflects the tendency for the opposite-sex allied children to appear more willing to accept blame for doing something wrong. Finding 4 indicates that the opposite-sex allied children appear to rank as "passive and to give in easily with adults," when compared to the same-sex allied children who rank as appearing more "aggressive and demanding with adults." Finding 5 indicates greater physical activity on the part of the opposite-sex allied children.

Although the data are presented more as a research curiosity than as definitive of any specific theory, one conclusion might tentatively be made. The profile of the opposite-sex allied children appears to be one of physical activity, personal blame-acceptance, conformity, and passivity with adults. The same-sex allies appear physically inactive, blame-avoiding, nonconforming, and aggressive with adults. Perhaps it may be concluded that the

TABLE 1

RELATED FACTORS IN FELT PARENTAL ALLIANCE

FINDING	CHI SQUARE
When compared to same sex allied children, the opposite-sex allied children:	
1. tend to win the conflict more	4.93
2. appear more conforming	4.00
3. appear more blame-accepting	4.00
4. appear more passive with adults	4.00
5. appear more physically active	3.79

Note.—All chi squares are significant at less than the 5 per cent level. For Finding 1, $N = 24$; for all other findings, $N = 37$.

[1] The research findings are the result of an investigation on the psychodynamic aspects of diabetes in children, sponsored by the Diabetes Camp, Home and Hospital Fund, and the Joslin Clinic, Boston.

opposite-sex identifiers are a little more satisfactorily adjusted to the characteristic childhood role. This hypothesis may be supported partially by the opposite-sex allied children's more frequent success in the symbolic familial conflict.

REFERENCES

1. BELLAK, L., & BELLAK, SONYA S. *Children's Apperception Test*. New York: C. P. A. Co., P. O. Box 42, Gracie Station, 1949.
2. TAT NEWSLETTER, *J. proj. Tech.*, 1951, **15**, (4).

Received May 27, 1953.

AUTOKINETIC JUDGMENTS AS A MEASURE OF INFLUENCE[1]

HARRIET B. LINTON

Yale University

SINCE Sherif (5) made the valuable discovery that the instability of judgments of autokinetic movement creates a useful experimental situation for the study of social influence, this technique has been used in a large number of experiments. In many cases, change of autokinetic judgments under social influence has been used as a criterion for other kinds of behavior, such as test responses, which purport to measure suggestibility (3, 6). A survey of the literature, however, reveals not only that the results are often negative, which is to be expected, but also that different experiments dealing with similar problems often produce contradictory results.

It seems likely that one reason for the equivocal results often obtained by using the autokinetic situation lies in the measures that are used. In most experiments a given degree of "influence" is applied to all *S*s; e.g., a planted series of judgments averaging five inches above the *S*'s initial norm. The most frequent measure used is the mean number of inches (or other unit) that each *S*'s judgments increase under the influence of the planted series, with this measure used in subsequent manipulations of the data as a continuous distribution. The method proposed here differs in two respects: (*a*) a measure of statistical significance (the *t* ratio) is used for each *S* rather than the actual magnitude of increase in judgments; and (*b*) the sample is dichotomized.

1. In most experimental work, psychologists are aware that a large difference between two groups of measures does not in itself give reliable evidence of a real difference. The significance of a difference likewise depends on the variability of the measures, a fact that is taken into account by the statistical techniques for evaluating the significance of a difference. In the autokinetic experiment, an increase of 1.5 inches from one series of 20 judgments to another cannot have the same interpretation for an *S* whose judgments have a sigma of 5.0 inches in each series that it has for an *S* whose judgments have a sigma of 0.5 inches. The estimates of the latter have clearly shifted in range, while those of the former have not necessarily done so.

A phenomenological basis for choosing the most valid measure can be sought in the statements of *S*s about the nature of influence. In a series of experiments (2; 3 [wherein the procedure is described]), 14 *S*s who felt that they had been influenced by a confederate's responses were asked what the increase in their judgments represented. Nine felt that the movement of the light had not increased, but stated that the confederate had influenced them to change their scale; four felt that the light moved more, but not as much as the increase in their judgments, and also reported a change in scale; only one reported an increase in movement without a change in scale. In other words, all but one of the *S*s who felt that they were influenced spontaneously stated that the influence caused them to change the scale with which they estimated movement. This evidence supports the use of a *t* ratio rather than the actual magnitude of the increase, since it is only by using a measure of statistical significance that a change in the range of judgments can be identified.

In the study reported, the criterion for influence was a *t* of at least 4.50, derived from a comparison of an initial norm based on 10 trials and the mean of a series of 40 "influence trials." The actual criterion used in any particular experiment naturally depends on the number of trials on which the measure is based. It should be noted, however, that statistical significance in itself is not an adequate criterion for change of scale. Many *S*s who showed none of the other attributes which were found to be typical of *S*s classified as "influenced" showed a statistically significant increase in judgments. The criterion should be high enough that it can be assumed to represent a genuine change in the scale used by the *S*.

2. The practice of dichotomizing has both theoretical and practical support. In the experiment cited above, *S*s were first divided into three groups:

[1] This study was done while the writer was a Public Health Research Fellow of the National Institute of Mental Health.

those with $t \geq 4.50$, those with $t < 4.50$ and ≥ 2.00 (significant at 5% level), and those with $t < 2.00$ or an actual decrease in judgments. On the basis of interviews with *Ss* about their reactions to the autokinetic situation, to other experimental measures, and to personality tests, it was found that the two lower groups did not differ in any discernible respect, but that the first group was clearly different from the other two. It was then decided that the criterion of $t \geq 4.50$ represented, under existing conditions, a change of scale, and that the variability within the two resulting groups probably resulted from other causes. The practice of dichotomizing is further supported by evidence in other experiments which indicates that influence from another person tends to be reacted to in an all-or-none fashion, and hence that a bimodal rather than a normal distribution of responses results (4).

There is still the question of why one-third of the present sample showed a statistically significant increase in their judgments, but did not show the other characteristics of *Ss* who were influenced. The explanation is probably to be found in changes in the autokinetic phenomenon itself. Graybiel and Clark (1) have found that over a long series of judgments *Ss* show consistent changes in their reports, some *Ss* increasing consistently, while others decrease. It is likely that the reliable increases shown by *Ss* in the middle group reflect the same phenomenon (a few *Ss* in the lowest group showed reliable decreases in judgments).

3. When different measures of change of judgments, based on the same raw data, are used to compute correlations with other variables, they lead to different results. The change in autokinetic judgments was correlated with 14 dichotomous Rorschach-test variables that were considered to reflect aspects of personality relevant to suggestibility (cf. 3 [wherein eleven variables are described]). To test the usefulness of different techniques for measuring autokinetic change, the Rorschach variables were correlated with three autokinetic measures: (*a*) biserial r with mean number of inches each *S* increased his estimates (the measure used most often in other studies), (*b*) biserial r with each *S*'s t ratio, and (*c*) tetrachoric r with the sample dichotomized at $t = 4.50$.

Using biserial r with mean increase in judgments produced only three correlations of .30 or more, all less than .50 (median = .13); biserial r with the t ratio produced nine correlations of .30 or more, two of them higher than .50 (median = .37); tetrachoric r produced eleven correlations of .30 or more, seven of them higher than .50 (median = .51). Another way of comparing the three methods is in terms of the statistical significance of the obtained correlations: of the biserial r's with mean increase, four were significant at the 10% level, three at the 5% level, and one at the 1% level; of the biseral r's with the t ratio, nine reached the 10% level, eight reached the 5% level, and four were at the 1% level; of the tetrachoric r's, eleven were significant at the 10% level, nine at the 5% level, and five at the 1% level. It should be noted that the criteria for significance are more stringent for tetrachoric r than for biserial r.

Comparison of the three methods for evaluating autokinetic change shows, then, that higher correlations are obtained by using the t ratio than by using the actual increase in judgments, and that these correlations become higher when the t ratio is dichotomized. The possibility that the results are an artifact of the particular variables on which the correlations are based is reduced by the fact that the decision to use the dichotomized t ratio was made on the basis of other experimental measures before the Rorschach data were obtained.

Personality test measures, then, are highly related to change of autokinetic judgments when a measure of significance is used and dichotomized, while the relationships become much lower when the magnitude of the actual increase in judgments is used. This suggests that, in the latter instance, potential correlations are obscured by extraneous variability, resulting from factors with little or no relationship to the personality factors being considered. It is suggested that the extraneous variability is a function of variations in actual autokinetic perception, while the process known as "suggestibility" affects the scale with which the *S* judges movement (i.e., modifies his conceptual and verbal behavior) without necessarily modifying his actual perception.

CONCLUSION

It seems likely that in experiments using the autokinetic phenomenon to study suggestion the data reflect two different factors: (*a*) the actual perception of autokinetic movement, which is subject to variability, but is probably influenced little, if at all, by social suggestion, and (*b*) a judgment process, which can often be influenced by social suggestion. The type of measure which has been here suggested attempts to evaluate changes in the judgment process, and to reduce the effect of perceptual changes on the measure used. This seems to be a desirable goal in experiments that attempt to use the autokinetic situation to study social processes. It seems likely that many experiments that have reported negative results could yield positive results if the data were analyzed in this way.

REFERENCES

1. GRAYBIEL, A., & CLARK, B. The autokinetic illusion and its significance in night flying. *J. Aviat. Med.*, 1945, **16**, 111–151.
2. LINTON, HARRIET B. Relations between mode of per-

ception and the tendency to conform. Unpublished doctor's dissertation, Yale Univer., 1952.
3. LINTON, HARRIET B. Rorschach correlates of response to suggestion. *J. abnorm. soc. Psychol.* 1954, **49**, 75–83.
4. MURPHY, G. *Personality, a biosocial approach to origins and structure.* New York: Harper, 1947.
5. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.
6. STEISEL, I. M. The Rorschach test and suggestibility. *J. abnorm. soc. Psychol.*, 1952, **47**, 607–614.

Received May 14, 1953.

SOCIAL RESEARCH AND THE PROBLEM OF VALUES: A REPLY

ELSE FRENKEL-BRUNSWIK

University of California

IN HIS paper, "How Neurotic is the Authoritarian?" Masling (8) raises the objection that undue bias entered our description of the authoritarian personality (1). Since I consider this problem to be very crucial and since the issue of undue influence of the authors' values in *The Authoritarian Personality* has been repeatedly raised, I accepted the suggestion of the editor of this journal and of Masling to discuss the latter's paper. In doing so, I shall start with the more concrete issues that Masling raises and then proceed to a more theoretical consideration of the problem involved.

One of the major points in Masling's paper is that the authoritarian personality is described in such "nasty" terms as submissiveness to authority, aggressiveness against the weak, superstition, stereotypy in thinking, conventionality, or cynicism. All these characterizations are said to presuppose maladjustment and pathological traits in the authoritarian personality, whereas, in reality, there is no evidence that the authoritarian is more neurotic than the equalitarian. Masling cites four independent studies that show no relation between authoritarianism and mental health, a result which he assumes to contradict the major assumption of our book.

In reply it must first be pointed out that there are explicit statements in the main body of our book to the effect that there is no direct relationship between authoritarianism and mental health. On the basis of my analysis of the interviews (1, ch. IX–XIII), we had even come to the conclusion that the use of projection mechanisms as manifested in prejudice helps rather than hampers the authoritarian in preserving his mental balance, especially in view of the fact that in this case defense mechanisms act through channels which are partially socialized (p. 480). The following quotation from the same context will complement this by showing that the opposite, equalitarian personality type is by no means depicted as an ideal of mental health:

> While for the high scorer [high on explicit ethnocentrism] possible loss of energy is connected with his tendency toward rigid repression, the low scorer is apt to waste energy by indulging in often unfruitful introspection and by placing the blame for mishaps too much upon himself. In contrast to the high scorer's tendency toward externalization, the typical low scorer is prone to internalize in an excessive manner, and this in turn may lead to open anxiety, feelings of guilt, and other neurotic features (p. 475).

In discussing the greater readiness of the equalitarian personality to face his conflicts openly, we stress that "temporarily, however, such frankness may well lead to increased anxieties and depressions."

Neurotic features are generally fairly common among low scorers on ethnocentrism, but they are especially frequent in a subtype that we have called the "rigid low." To quote from our book: "It must be reiterated . . . that there is a distinct type among low scorers in whom the liberal ideology becomes a cliché and who at the same time shows rigidity in his personality make-up" (p. 481). This also is the subtype who is somewhat predisposed toward certain forms of totalitarianism even though these may be of a variety opposite to Nazism or other rightist ideologies.

> With [them] . . . the struggle for minority rights . . . contained features of compulsiveness, even of paranoid obsession. . . With respect to many of our variables, especially rigidity and "total" thinking, [they] could hardly be distinguished from some of our high extremes. . . . The latter kind of low scorers are definitely disposed towards totalitarianism in their thinking; what is accidental up to a certain degree is the particular brand of ideological world formula that they chance to come into contact with (p. 772).

Instead of a continuum we here have partial curvilinearity, with some overlap between the rigid low scorer and the high scorer; curvilinear relations are generally frequent in personality studies and may be subsumed under the heading of "closeness of opposites" (4).

The findings of Maria Levinson, partly based on the Minnesota Multiphasic Personality Inventory, to the effect that there is no clear-cut relation between authoritarianism and mental health (1, ch. XXII) are thus not in disagreement with the rest of our book. It seems that our rather dynamically conceived syndromes cut across the classifications prevalent in psychiatry, which are also used in scoring the MMPI. These classifications may be too undifferentiated, or rather too oblique, for our purposes, covering as they do a variety of characteristics which very likely are inhomogeneous as far as our specific syndromes and their psychological dynamics are concerned.

In a more recent paper Gough (6) even found some relationships between authoritarianism and the clinical scales of the MMPI, such as those for psychopathic deviate or for schizophrenia. More marked relationships were found between authoritarianism and personality clusters based on an analysis of individual items. Gough concludes that authoritarians tend to be resentful, harrassed, embittered, bewildered, and to display certain paranoid traits.

Let us now turn to the results of Knutson as reported by Masling. There is an absence of relationship between the score on Sanford's (9) authoritarianism scale, which is a modification of our F scale, and the score on a "Personality Security Scale." Such contradictions can usually be reconciled by reference to different levels of personality structure. We found in the intensive clinical interviews with authoritarians that while on the surface they often profess self-assurance, on further probing they turn out to be rather insecure. In a previous study on adolescents (2, p. 243), I was able to demonstrate that both the most secure and the most insecure individuals tend to score high on items that constitute a rather direct inquiry into feelings of security. Extreme overt self-satisfaction, far from being a dependable sign of adjustment, often just indicates that repressive mechanisms are at work. The curvilinear relationship obtained between self-reported adjustment and objective adjustment as rated by clinicians (2) allows us to expect that authoritarian individuals will tend to score either quite high, or else quite low, on a security scale, depending upon the directness of the items. One of our most significant findings with authoritarian individuals was that they tend to have an explicit image of themselves which is in contradiction to the one revealed by a more searching depth evaluation of the interviews and projective techniques. This is only one of the many inconsistencies in the authoritarian personality. Most of these inconsistencies are the result of an inherent discrepancy between surface and depth. This also holds for the principle of the closeness of opposites as it appears in the coexistence of rigidity and extreme fluidity, of overcautiousness and a tendency toward short cuts to action, of submission to and resentment of authority, of emphasis on control and an uncontrolled yielding to impulses.

The qualifications that have just been made in connection with overt security scales also apply to the other structured, or semistructured, tests referred to by Masling. The elicited answers may have originated in the deeper layers; or they may have come from the surface layers, being largely determined by the compensatory defense mechanisms which are so typical of the authoritarian personality. Relationships would thus tend to be obscured. The clinical interview, with its deeper probing and the opportunity it gives the subject for spontaneous formulations of the issues in question, provides us with sufficient cues for differentiation between the surface layers and the deeper layers. Probings of this kind may be able to give the key to the fact that, in a study on Navy recruits reported by Masling, no relationship could be found between authoritarianism scores and hospitalization. Hospitalization is a behavioral end effect and thus a surface phenomenon that may have a variety of dynamic causes. In describing the personality patterns of the hospitalized recruits, Masling points out that they generally tended to display immature behavior and emotional instability in forms frequently leading to maladjustment. We can well imagine that in the nonauthoritarian a given Navy or Army situation could induce manifestations of anxiety and instability along the line of open resentment against authority figures and against the overstructured situation; in the authoritarian, on the other hand, it may be mainly the situational change as such and the temporary loss of the clear-cut structure provided by the family which is the source of maladjustment.

It was primarily the above-noted complexity of the relationships between the underlying states and the overt behavioral manifestations that prompted me in another context (5) to advocate the use of a different vocabulary for underlying aggression, anxiety, and so forth, on the one hand, and manifest aggression, anxiety, and so forth, on the other. The use of the same term for both layers creates the illusion that we are dealing with the same set of facts; in reality, however, there are —at least—two sets. These vary independently to a certain degree, and the curvilinear character of the relationships referred to above further bespeaks the complexity of the patterns involved.

Let us now turn to what is ostensibly Masling's main objection. This is to a certain degree independent of the problem of the relationship between authoritarianism and mental health. It concerns the somewhat broader observation that

evaluative terms are used at all in our description of the authoritarian personality. Masling is certainly right in taking for granted that such terms as prejudice, aggression, and rigidity carry some negative connotations. Statements containing these and related terms are by no means necessarily value judgments in the technical sense, however. The question arises here whether or not as a matter of principle terms with evaluative connotations can or should be avoided at all in the social sciences. Almost all trait names are evaluative, imbued as they are with functional versus nonfunctional or malfunctional connotations. It would be extremely awkward to list each time a series of sober operational specifications instead of using the emotionally somewhat loaded terms of the vernacular or of ordinary personality psychology. And even so, instead of saying that authoritarians are rigid, we would have to say that they tend to cling to the familiar and definite, that their approach stresses detail and is generally stimulus-bound, that they have difficulty in changing their attitude or set, that they are apt to search for absolute solutions, that most of their actions stem from the need to protect themselves rather than being free and spontaneous, and that their feelings and thoughts are highly channelized, making them relatively inaccessible to new experience; obviously this attempt to eliminate the general term by giving behavioral specifications would not rid us of evaluative involvement.

The reason is obvious; there are richer and poorer, more efficient and less efficient orientations or ways of dealing with reality, and as scientists we are not entitled to obscure or to circumvent this fact. Yet it is important to remember that statements asserting the presence of such survival-relevant traits as rigidity, conventionality, superstition, or punitiveness in a person or group can fully be translated into more descriptive statements about ways of dealing with reality, and that such statements do not constitute value judgments in the sense of an expression of preference on the part of the investigator. In keeping with this, such terms as "aggression against the weak"—to which Masling objects as evaluative—acquire a different flavor as soon as we remember that "aggression" designates a certain way of functioning in a descriptive manner or as a dispositional term, and thus is a legitimate part of the behavioral vocabulary, and that "weak" is defined by a certain objective locus in the social hierarchy. Similarly, "superstitious" designates nothing but the presence of a belief in supernatural explanations of natural phenomena. The sting that is bound to remain in any reference to such attitudes is given by the fact that they are in general ill suited for coping adequately with the complexities of social and cognitive problems. They tend to lead to consequences that are detrimental to society, to others, and often to their holder himself; but this tendency is empirically verifiable like any other factual trend.

This situation is by no means unique to psychology. In chemistry such originally rather emotive terms as "noble" versus "base" have for long periods of time been used in the purely scientific characterization of certain properties of metals that are relevant to their own "survival," such as their ability to resist corrosion.

It must be granted, nonetheless, that as long as we allow ourselves to become enmeshed with adaptive versus maladaptive ways in the ambiguous sphere of emotions and social attitudes, there is danger that the values of the researcher will enter the field and that the facts will be overinterpreted in the direction of his prescientific value system. Awareness of this danger led me to undertake perceptual experiments with ethnocentric and nonethnocentric adolescents in which the reality, presented in the form of a story or a perceptual stimulus, was known beyond doubt, and thus the reality-adequate versus narrow or distortive handling of this reality could be ascertained in an objective, indisputable manner (4). The results of this study as well as those of subsequent independent investigations have confirmed that the authoritarian personality tends to a more limited and narrow approach to reality, and at the same time distorts it to a greater degree than the equalitarian.

As we have stressed again and again in our book, these nonoptimal forms of approaching reality are by no means to be taken for any gross maladjustment or mental deviation. In fact, as we have concluded on the basis of our interviews, the authoritarian may, and frequently does, function extremely well in most routine contexts of life; he probably even betters the equalitarian under certain narrowly circumscribed conditions, such as when the culture provides socialization for the mechanisms involved. Nazi Germany is an excellent example of a people functioning well within a certain given framework. This remains true in spite of the fact that in due time the reality-inappropriate expectations of a final victory of the master race and the projective and thus distortive view of the democracies as effeminate societies resulted in self-destruction. Catastrophies like this have confirmed us in our assumption that the authoritarian syndrome possesses a reduced survival value in the long run, adaptive as it may be for the short haul.

With all this in mind, we are still far from deny-

ing that the authors of *The Authoritarian Personality* brought with them, as would most citizens in general and most social scientists in particular, a definite set of values concerning the very issues that were under investigation. In our Introduction we explicitly recognized that our own values differ considerably from those expressed by our authoritarian subjects. In holding such preconceptions, social scientists differ from other scientists only in degree, not in kind. Choice of the field of study, and of the specific problems of research within this field, seem to be issues the social scientist shares with all other scientists. These choices are largely based on prescientific value decisions; there can be little doubt that they are determined at least in part by social background, personality, and other factors in the biography of the individual.

Ideally, the only value system that should be involved while engaging in scientific work is what Spranger has called "theoretical values." More aptly, it may be characterized as the value system of scientific objectivity. However, as we witness even in the most abstract and logically rigorous theoretical systems in psychology—say, Hull's postulational theory of learning—or in physics, science as an actual human undertaking has always been full of subjective preconceptions. These preconceptions are often of an esthetic or formalistic character, containing a strong element of bias or distortion. Those denying the existence of such biases in their own work in most cases turn out just not to have made them sufficiently explicit to themselves. In this respect, the social scientist is as much a child of his period with its varying climates of conceptualization as are his colleagues in other areas of psychology. Over and beyond this, however, there is probably no way for the social scientist to fully escape infiltration of his outlook by the sociopolitical climate of opinion in which he lives, no matter how hard he may try. An analysis of the subtle effects of the attitudes involved would probably be as difficult as an analysis of the fact that, say, psychologists in some countries have tended to become Gestalt psychologists while in others they became behaviorists. It is one of the foremost tasks facing the historian of science and culture, if not the social scientist, to make explicit and to trace the influences of the biases inherent in declared or tacit allegiance to any of these so-called "points of view."

The only important question in our context is whether or not the preconceived social and cognitive points of view were able to interfere with scientific rigor, that is, with the objective treatment of the specific problem under investigation. So far as our study of authoritarianism is concerned we must show that the personality relationships that were presented as empirically established are not merely the expression of wishfulness on our part. Several arguments may be brought forth to prove our original contentions.

First, we tried to make use of the most objective and quantitative techniques available in the social sciences today. In the evaluation of the interviews we tried to reduce bias to a minimum by the adoption of a "blind" scoring procedure and by employing a number of independent raters. A confirmation of the interview interpretations is given by the fact that the reactions on our F scale, purporting to measure in an indirect, projective way the personality traits associated with authoritarianism, show that our authoritarian subjects tended to present themselves precisely as we expected them to do according to our theory based on the more interpretative approach.

Second, as mentioned above, the reactions of authoritarian children to such clearly defined stimuli as perceptual configurations, stories, or cognitive tasks show tendencies of rigidity, distortion of reality, and intolerance of ambiguity similar to those found in the more controversial social and emotional area (4).

Third, in a separate California study on prejudice in children and adolescents (3), it was demonstrated even more directly that many of the tendencies that were found to be accompaniments of authoritarianism and ethnic prejudice at all age levels—such as superstition, intolerance of ambiguity, tendency to dichotomize, admiration of strength—are at the same time more frequent in chronologically earlier stages of development. In other words, the direction of the difference between prejudiced and unprejudiced subjects in general, with age held constant, is the same as that between the younger and the older child or adolescent. The assertion that the authoritarian personality is one of relative immaturity is meant in this strictly descriptive rather than in an evaluative sense.

Fourth, one of the most decisive arguments may be derived from a comparison of the results on the same topic of social scientists holding different prescientific preferences. From the purely argumentative point of view it is unfortunate that research concerning war and prejudice is in this country engaged in predominantly by those who do not like war or prejudice. Under this same aspect we must welcome the fact that we have available in the writings of Jaensch (7) a very articulate description of the ideal personality type from the Nazi point of view. While my own work on the perceptual reactions of authoritarian and nonauthoritarian children was in many ways in-

fluenced by the German typologists, among whom Jaensch had long been a prominent figure, his more specific writings on the Nazi ideal of personality did not reach me until most of the work had been done. Jaensch finds, as we do, the Nazi personality rigidly stimulus-bound. But unlike ourselves, Jaensch evaluates this rigidity as a positive trait and looks down upon the intellect in particular, and upon "adaptability" in general, as expressions of degeneration and morbidity. In this and many other ways a comparison between his and our findings reveals that investigators with opposite evaluative attitudes toward certain social phenomena may well reach the same conclusions concerning the factual aspects of the phenomena concerned.

Situations such as this, along with logical considerations, led Stevenson (11) to suggest a clear-cut distinction between value judgment, on the one hand, and scientific or factual judgment, on the other. A similar point of view has been taken by most logical positivists. Ethical judgments, that is, statements containing such terms as "good," "desirable," "ought to," etc., are considered to be emotive expressions in which preferences are stated. Stevenson is especially impressed by the fact that there can be ethical disagreement—"disagreement in attitude"—without a disagreement about facts, that is, without "disagreement in belief." Thus, he points out that two persons can fully accept the same facts and still differ in their attitudes toward the facts. No other method of convincing the other person is available in such cases except the method of persuasion. Though often our attitudes are based on facts, neither facts nor scientific methods can be granted a definitive role in a "normative science" such as ethics.

Another group of value theories may be subsumed under the naturalistic point of view in that statements about values are not considered to be different from empirical statements but assumed to be open to verification in the same manner as other scientific statements. According to this view, moral predicates can be restated in nonmoral predicates, such as that of "better" conduct as one of "greater complexity"; in general, restatement of moral judgments takes place in terms of consequences. Thus, the certainty of moral statements is seen as neither greater nor less than that of scientific statements.

In criticizing the naturalistic point of view, an earlier representative of positivist ethics, Schlick, states: "A fundamental error lies at the basis of the whole attempt; it consists in seeking value distinctions in the objective facts themselves, without reference to the acts of preference and selection through which alone values come into the world" (10, p. 104).

The logical positivist point of view concerning values has been criticized mainly on the ground that this position sees in value statements purely arbitrary expressions of emotions and preferences. Suffice it to say here that one may counter this argument by pointing out that although ethical norms must be differentiated from facts, facts play a part in all value statements. Furthermore, individuals and cultures do, on the whole, not differ widely with respect to what are considered the ultimate ethical goals, so that the arbitrariness in preferences is more a matter of means. Even Jaensch would doubtless agree with us on many values, such as the desirability of health—a factor which, by the way, he stresses a great deal.

Although we follow Stevenson to some extent in his view of the subjective nature of some of our ultimate value judgments or attitudes, we find his distinction between factual and value statements in many respects overdrawn. To begin with, statements and agreements about facts are not as clear-cut and uncontroversial as meets the eye, due largely to the preconceptions which enter our view of the facts from the very beginning. Even in the most "objective" sciences, the so-called theoretical structure, although guided by experience, is in the main a product of our imagination and is connected only loosely, by so-called interpretative links, with the observational data, thus allowing for alternative explanations of the same facts. Furthermore, in many cases an alleged value statement turns out to be fully translatable into statements about facts. In any event, value judgments are at least partially based upon fact, a circumstance which Stevenson also recognizes.

For all these reasons, the knowledge accumulated in the social sciences may help us to make a choice between alternative value systems. In renouncing any metaphysical or absolute position, we do not need to go to the other extreme of utter relativism. Although the social scientist, as a scientist, cannot make the ultimate choice for mankind, his function is to throw as much light as possible on the implications involved in existing value systems, and to make explicit all the ramifications inherent in the options. In our work on the authoritarian personality we tried to establish a network of empirical relationships between the holding of certain social attitudes, such as ethnocentrism, and a host of seemingly more neutral personality traits. Even though statements about personality traits almost always carry a certain evaluative reference as to the immediate or long-range adjustment of the individual or society, they are in themselves cognitive statements in that they refer to relations between empirical facts.

It was our hope that our work would increase our theoretical understanding of both personality

and society, and that it would add to the factual basis for such value decisions as to whether or not we should combat authoritarianism, be it one of the right or one of the left. But only the direct designation of authoritarianism as desirable or undesirable would constitute a value judgment in the logical sense of the word. Only the direct expression of such preferences, or the distortion of fact under the indirect influence of such judgment, would go beyond, or interfere with, scientific procedure. I have tried to show that we have steered clear of either of these impasses.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. FRENKEL-BRUNSWIK, ELSE. Motivation and behavior. *Genet. psychol. Monogr.*, 1942, **26**, 121–265.
3. FRENKEL-BRUNSWIK, ELSE. A study of prejudice in children. *Hum. Relat.*, 1948, **1**, 295–306.
4. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and cognitive personality variable. *J. Pers.*, 1949, **18**, 108–143.
5. FRENKEL-BRUNSWIK, ELSE. Psychoanalysis and the unity of science. *Proc. Amer. Acad. Arts Sci.*, 1954, **80**, No. 4, 271–350.
6. GOUGH, H. Studies of social intolerance: II. A personality scale for anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 247–255.
7. JAENSCH, E. R. *Der Gegentypus*. Leipzig: Barth, 1938.
8. MASLING, J. M. How neurotic is the authoritarian? *J. abnorm. soc. Psychol.*, 1954, **49**, 316–318.
9. SANFORD, F. H. *Authoritarianism and leadership: A study of the follower's orientation to authority*. Philadelphia: Institute for Research in Human Relations, 1950.
10. SCHLICK, M. *Problems of ethics*. (Trans. by D. Rynin.) New York: Prentice-Hall, 1939.
11. STEVENSON, C. L., *Ethics and language*. New Haven: Yale Univer. Press, 1945.

Received February 15, 1954.

ADMINISTRATION OF THE F SCALE TO A SAMPLE OF GERMANS[1]

THOMAS S. COHN AND HENRY CARSCH

Wayne University *Yale University*

The F scale has been administered to a wide variety of groups with the implication that differences in mean scores reflected differences in deeplying, potentially fascist personality trends established before the age of five. This report makes available data on the F scale obtained from a sample living in Germany. The scale was translated into German[2] and administered to 140 workers in a cosmetics factory in 1952. The mean F score was 5.26, *SD* of .86. This score is higher than the mean score for any group thus far reported in the literature ($p > .01$). Thus, the mean F score for San Quentin prisoners was 4.73, *SD* of .86; for working class men, 4.19, *SD* of 1.18; for professional women, 3.43, *SD* of .86; etc. (1). If one makes the assumption that these workers are "fascistic" personalities, then the data appear to support the validity of the F scale.

In studies using American *S*s, F scores are negatively correlated with educational level (2, 3). This relationship holds for our German sample. The mean F score of the 117 *S*s that had attended Volkschule or Volkschule and Mittelschule was 5.40, *SD* of .86, while the mean F score of the 23 *S*s that had attended Hochschule was 4.57, *SD* of .85. The difference between means is significant ($p < .01$). This finding indicates the possibility that the F-scale score is partly a result of educational forces operating after the age of five.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality* New York: Harper, 1950.
2. COHN, T. S. Factors related to scores on the F scale. Unpublished Ph.D. thesis, Univer. of Michigan, 1953.
3. LASKY, J. J. Relations of experimentally induced changes in ethnocentric statements to underlying personality variables. Unpublished Ph.D. thesis, Univer. of Michigan, 1950.

Received March 6, 1953.

[1] This study was carried out while the senior author was a member of the staff and the junior author a student at the University of California, Santa Barbara College.

[2] The authors would like to express their appreciation to Dr. Rolf Linn of the University of California, Santa Barbara College, for his aid in translating the F scale into German.

NOTES ON THE WESLEY RIGIDITY SCALE: THE DEVELOPMENT OF A SHORT FORM[1]

SEYMOUR L. ZELEN AND EUGENE E. LEVITT

State University of Iowa

In a recent article, Wesley (5) has reported the development of a paper-and-pencil scale for the measurement of rigidity. The scale consists of 50 statements with which the subject indicates agreement or disagreement. A 41-item form of the scale was also developed and has been used in two researches which cast serious doubt on the validity of the scale. Katz (2) found that it was related to only one of 11 measures of rigidity, the E scale. Moldawsky (3) devised an ingenious experiment in which a measure of rigidity was obtained from spontaneous behavior. This measure was not related to scores on the Wesley scale.

This paper is not concerned directly with the validity of the Wesley scale but rather with certain aspects of the scale as a scale. Examination of the scale reveals what appears to be considerable overlapping of items, for example, "I am a methodical person in whatever I do," and "People who go about their work methodically are almost always the most successful," and finally, "It is the slow, steady worker who usually accomplishes the most in the end." There are nine sets of such potential identities, totaling 26 items. In addition, there are four sets, taking in 11 items, which seem to be absolute opposites, like "I prefer to stop and think before I act even on trifling matters," and "I have taken a good many courses on the spur of the moment."

There is no reason to believe that the addition of an identical item or an item which is identically opposite enhances the efficacy of a scale. A test composed of many such items can hardly be considered a true scale. Whether or not the items are actually overlapping is best determined by empirical investigation.

From Wesley's original group of items (4), the authors selected 12 that did not seem to overlap and which they consider to be satisfactory for a rigidity measure.[2]

1. I am often the last one to give up trying to do a thing.
2. There is usually only one best way to solve most problems.
11. I dislike to change my plans in the midst of an undertaking.
12. I never miss going to church.
14. I would like a position which requires frequent changes from one kind of task to another.
20. I do not enjoy having to adapt myself to new and unusual situations.[3]
28. My interests tend to change quickly.
40. I am always on the lookout for different ways of doing something.
42. I always finish tasks I start, even if they are not very important.
44. When I have undertaken a task, I find it difficult to set it aside, even for a short time.
60. I like to surprise my friends by unexpected actions.
62. I find it difficult to change my way of doing something even though it may not be successful.

The test-retest reliability of this short scale was .74 for an *N* of 31 undergraduates. It correlated .73 with the 41-item Wesley scale for 123 subjects. In two groups totaling 87 subjects, the average correlation between the short scale and the California E scale was .38 with a standard error of .11.

These data on the 12-item form compare very favorably with the 41-item Wesley scale. Reliability of the 41-item form is not reported anywhere by Wesley, but the authors computed Kuder-Richardson reliabilities for Wesley's data (4) and for those of Katz (2) and Moldawsky (3). The coefficients are, respectively, .68 for 284 subjects, .59 for 72 subjects, and .67 for 453 subjects. According to Froelich (1), the Kuder-Richardson should not underestimate the split-half reliability by more than .06. We may assume, then, that split-half coefficients for the 41-item scale would not exceed .74. Since the split-half is commonly considered to be a best estimate of the test-retest reliability, we may say that the 12-item form is as reliable as the 41-item or the 50-item forms.

Katz (2) reported a correlation of .26 between the 41-item scale and the California E scale. This correlation is significant at about the .03 level. The comparable correlation obtained with the short form is .38, which is significant well below the .01 level. Apart from Wesley's (5) results with the Wisconsin Card Sorting Test (which are not above criticism), the E scale is the sole potential criterion to which the Wesley scale has been found

[1] This investigation was supported by Research Grant MH-301 from the National Institute of Mental Health of the National Institutes of Health, U. S. Public Health Service. This grant supports the Preventive Psychiatry Project which operates within the Child Welfare Research Station of the University of Iowa. The research was conducted directly under the auspices of the Preventive Psychiatry Project.

[2] The numbers are the same as in Wesley's original form (4).

[3] This item is phrased positively in the original form.

to be related. In this respect, too, the 12-item test is at least as good as the longer form.

The distribution of scores obtained with the short form is almost identical with those obtained with the long form. The coefficient of variation for the distribution of scores for the 123 subjects on the short form is 26 (a mean of 6.46 and a sigma of 1.69). The V's derived from the data of Wesley, Katz, and Moldawsky are respectively 24, 26, and 28.

It is quite true, of course, that the .73 relationship between the short and long forms is influenced by the presence of the short-form items within the long form. This circumstance, however, is immaterial for our purpose, which is merely to demonstrate that at the present stage of development, the short form may be substituted for the long form with savings of time and effort. The correlation between the two forms is as high as the reliability of either form, indicating that to all intents and purposes, the 12-item scale and the 41-item scale are comparable forms of the same measure.

REFERENCES

1. FROELICH, G. J. A simple index of test reliability. *J. educ. Psychol.*, 1941, **32**, 381–385.
2. KATZ, A. A study of the relationships among several measures of rigidity. Unpublished doctor's dissertation, Univer. of Iowa, 1952.
3. MOLDAWSKY, S. An empirical validation of a rigidity scale against a criterion of rigidity in an interpersonal situation. Unpublished doctor's dissertation, Univer. of Iowa, 1951.
4. WESLEY, ELIZABETH. Perseverative behavior in a concept formation task as a function of manifest anxiety and rigidity and of punishment. Unpublished doctor's dissertation, Univer. of Iowa, 1950.
5. WESLEY, ELIZABETH. Perseverative behavior in a concept-formation task as a function of manifest anxiety and rigidity. *J. abnorm. soc. Psychol.*, 1953, **48**, 129–134.

Received February 4, 1954. Prior publication.

THE THERAPEUTIC COMMUNITY. A NEW TREATMENT METHOD IN PSYCHIATRY. By Maxwell Jones in collaboration with A. Baker, T. Freeman, J. Merry, B. A. Pomryn, J. Sandler, and Joy Tuxford. Foreword by Goodwin Watson. New York: Basic Books, 1953. Pp. xx + 186. Price $3.50.

This monograph by a British social psychiatrist and his collaborators describes an attempt to develop a new method for the vocational and psychological rehabilitation of those most difficult of all nonpsychotic psychiatric patients, the chronically unemployable neurotics, psychopaths, and character disorders. Sponsored by the Ministries of Health, Labour, and Pensions, an Industrial Neurosis Unit was established with a small staff of psychiatrists, psychologists, social workers, government "disablement resettlement" officers, occupational instructors, a research technician, and nurses. Patients with psychiatric disabilities who had particular employment problems were sent to this unit from around the country. The method of treatment that was developed is clearly presented in the text with reports on its psychiatric, psychological, vocational, and social aspects. Some interesting research designed to improve procedures, evaluate the effect of the method against certain controls, and analyze pertinent factors in the employability prognosis of the patients is included, making for a well-rounded, thought-provoking contribution to the modern vocational rehabilitation literature.

After several brief accounts of earlier units, the treatment method of this Industrial Neurosis Unit is described in detail. Shortly after receiving fairly intensive psychiatric and psychological evaluation, new patients are required to begin working regularly (four hours per day) at one of a number of semiskilled or unskilled jobs available in the hospital workshops or nearby industries. The aim is "to get them back to the habit of work." This work is closely supervised by the staff and is extensively supplemented by morning group discussions, some individual psychotherapy, and considerable emphasis on and opportunity for socialization, group communication, and hospital community responsibility. While the focus of concern is always vocational rehabilitation, the basic objective toward which the program is geared is that of providing for these asocial, embittered personalities a rich and rewarding experience in group living, in belongingness, and in the satisfactions of community effort and responsibility. To accomplish this objective, two major approaches were involved: (*a*) an imaginative and varied use of group psychotherapy for the patients and (*b*) a marked reorientation of the attitudes and role of the hospital staff. The closeness of living and working in this benevolent but fairly structured environment provided extensive material for the morning group psychotherapy sessions. Didactic lectures, opportunities for group catharsis, some quasi-analytic group methods, as well as role-playing, psychodrama, and free-expressive methods (drawing, for example) were all employed. The integration of these varied techniques of modern group dynamics seems especially skillfully handled and is well exemplified throughout, particularly in the rather dramatic excerpts presented in the appendix.

Perhaps even more unique in this study is the extensive concentration on developing *within the staff* an atmosphere and general orientation conducive to helping the patients on their road to socialization. The self-searching consideration of the various professional roles, the use of enlightened group dynamics techniques within the staff conferences are in many ways the most daring and refreshing feature of the entire project. In fact, the many practical problems of changing functions and staff attitudes which are clearly presented should prove especially provocative to readers from the various disciplines engaged in the vocational rehabilitation of the psychiatric patient. Translation of these British methods to the rehabilitation work in our own state and veterans hospitals, even with the considerable differences in the extent of centralized national control, should prove relatively feasible.

Regarded as a progress report on the first stages of a long-term effort, this monograph is indeed valuable. True, it is short on theory and the discussion of group dynamics, and the therapeutic process is somewhat superficial. Nevertheless, the manner in which ongoing research is integrated with the therapeutic process as an essential feature of the total program is gratifying, whatever the limitations of the actual data. The practical problems raised and the nature of the material have a social realism that extends far beyond the limited scope of many clinical investigations. Thus, a deeper note is struck when the caseworker doing the follow-up home visits comments on the significant lack of social cohesion she observed in many communities. Problems in marriage, employment, and juvenile delinquency are increasingly being solved by resort to agencies and external officials rather than by initial efforts within the community or even within the family structure itself. In the face of this broader weakness in the social scheme, the admittedly tentative results obtained by these investigators in their effort to

integrate these chronically asocial patients as productive members of the community are certainly encouraging.

JEROME L. SINGER
Franklin D. Roosevelt VA Hospital,
Montrose, New York

THE NATURE OF NONDIRECTIVE GROUP PSYCHOTHERAPY. AN EXPERIMENTAL INVESTIGATION. By Leon Gorlow, Erasmus L. Hoch, and Earl F. Telschow. Foreword by Nicholas Hobbs. New York: Bureau of Publications, Teachers College, Columbia University, 1952. Pp. viii + 143. Price $3.25.

This study was designed to investigate with quantitative precision various parameters of the process of nondirective group psychotherapy. Acknowledging the tentative, exploratory nature of the investigation, the authors outline a threefold objective: (*a*) analysis of the uniformity of group patterns of verbal behavior as a function of variables like time, differing membership, and leadership; (*b*) delineation of the therapeutic value of the group, per se; and (*c*) evaluation of the role of the group leader. To achieve these goals each of the investigators led a nondirective psychotherapy group consisting of graduate students who expressed a desire for help. There were 17 subjects divided among the three groups. They met twice weekly for approximately ten weeks. Prior to therapy, subjects were evaluated for adjustment by means of the Rorschach, modified Hildreth Self-rating checklist, and an incomplete sentences test. Following treatment the last two tests and the investigators' judgments were employed to divide clients into *most-* and *least-*profited groups.

After an intensive evaluation of the reliability of ratings based upon recorded protocols, a chapter by Hoch provides data relevant to the hypothesis that there exists a characteristic nondirective group psychotherapy process. Thus, protocol ratings for the three groups reveal that expression of positive emotion increases with the passage of time, while negative attitudes reach their peak midway in the therapy cycle and subsequently decline. A subprocess is suggested by the finding that while the *most-profited* clients show a decline in "poorest" client behavior (confusion, defensiveness, negative attitudes towards self and others), the *least-profited* subjects persist in such behavior throughout the treatment cycle. Gorlow's section on the clients' activity as therapists reveals that therapeutic behavior denoting permissiveness and acceptance increases, while evaluation and criticism decline. Subjects rated from their Rorschachs as less anxious or hostile are categorized by therapist behavior involving more positive and less negative categories than that shown by more anxious group members. Finally, Telschow's section provides data suggesting relative uniformity of leader behavior with 98 per cent of leaders' statements categorized as nondirective. Except for a trend suggesting that greatest client release is evoked by clarification of feeling, results in this section are largely negative. In all, 40 conclusions are derived from the data by the three investigators.

This book is impressive by virtue of the extensive analyses of a wealth of protocol material. Certainly the collaborators have demonstrated many ingenious methods for quantifying group therapy data and they have laid bare many parameters of the treatment process that merit consideration. Nevertheless, even with the authors' frequently reiterated words of caution concerning the study's limitations, one cannot help but be disappointed with the book. The absence of any control groups raises a very serious question as to whether effects like the temporal sequences reported are peculiar to nondirective or even to any psychotherapeutic group. Can we be sure that a discussion group on mental health might not reflect similar patterns over a period of 20 sessions? Similarly, the criteria for evaluation of the most- and least-profited groups are not made sufficiently explicit, and they appear to be inadequate. The self-ratings or sentence completion items may simply elicit the very same response that characterizes the self-attitude statements rated from the therapy protocols. This seems to imply somehow that people who begin to make positive statements about themselves and others after 20 sessions of group therapy have profited from the treatment. This is a rather naive or narrow view which could have been clarified by more explicit presentation of external criteria of personality change to avoid the possibility that what develops among some group members is merely a mutual self-admiration society. The small number of subjects imposes a serious limitation upon the statistical treatment. As a consequence many of the results are inconclusive rather than clearly negative, and the investigators several times are put in the position of accepting the null hypothesis with correlations of the order of .50 or p's in the area of .10. The many trends suggested by the data deserve a more adequate test for conclusive acceptance or rejection of a number of important experimental hypotheses. Finally, although the book is clearly written, it is extremely abstract and dehumanized. The absence of actual protocols, frequent citations, or excerpts leaves one confused as to what actually goes on in the nondirective group psychotherapy session. This deficiency not only impairs the interest value of the book but denies the reader an opportunity to evaluate for himself the relevance of various response categories and of the authors' descriptions of the therapeutic process. In summary, then, this

book is successful in providing interesting hypotheses and techniques for analyzing protocols. The limitations of its experimental design, however, prevent an evaluation of either the unique characteristics or therapeutic value of nondirective group psychotherapy.

JEROME L. SINGER
Franklin D. Roosevelt VA Hospital,
Montrose, New York

THREE MEN: AN EXPERIMENT IN THE BIOGRAPHY OF EMOTION. By Jean Evans. New York: Alfred Knopf, 1954. Pp. xviii + 298. Price $3.75.

The psychology teacher who would bring his students the raw materials of human living finds himself facing a singular dilemma. Scientific literature is replete with case studies setting forth the salient facts about the lives of various people. In this journal alone, in the decade 1939–1949, no less than 25 such studies appeared. However, nearly all available case studies, when considered as teaching materials, are found to have the serious drawback of including the writers' hypotheses and interpretations. This is no doubt desirable from the point of view of scientific communication. But what happens when you put such a document in the hands of a student? Usually he feels that there is little he can do with, or add to, the material. The author, he feels, knows more about the subject than he does, and has already set forth the most likely hypotheses concerning dynamic and causative factors in the subject's life and developed them more adequately than he, the student, could possibly do. So he is often reduced to reading the study and wondering, in a frustrated sort of way, at the wisdom of the author; he feels also, perhaps, that he will never reach that point of wisdom in his own scientific development.

What are needed here are the sheer, specific facts of a single life, unadorned by any interpretation, explanation, or elaboration. True, there must be selection. No one tome would be large enough to contain all the facts of even the most ordinary life. But what the student needs is a selection of those facts that mirror the life of the subject, as he lived it, in his own world. Schools of law and of business administration have pioneered in adapting the "case method" of instruction to the social sciences, and in developing case materials free of interpretation and explanation. Donham at Harvard and Berrien at Colgate have been leaders in this effort. A further drawback, for which the authors can hardly be blamed, characterizes most of the case materials so far published. The scientific or professional psychologist is rarely, if ever, an artist and craftsman in the field of letters. His presentation of the facts of a life story may achieve objectivity, but it seldom rises to that graceful and inspiring verisimilitude that distinguishes the work of a really good professional writer.

Jean Evans is a professional journalist and author whose work combines in a truly unusual way the sensitivity and felicity of expression of the artist with a passionate devotion to the ideal of what Allport has called "portraiture." "Walk through the slum section of any American city some evening. Pause at the poolrooms, the gyms, the dingy bars, the candy stores, and certain street corners where boys and young men gather. Any one of them might be a Johnny Rocco." Thus begins the life story of "Johnny Rocco," which first appeared in this journal in 1948. In 1950 the journal published Miss Evans' "portrait" of "William Miller," son of Danish immigrants to a small town in Pennsylvania. He had lost both parents by the time he was six months old. In the course of an unusual life, involving a variety of hardships and ranging over a wide geographical area, he lost and recovered his vision several times in ways and under circumstances that totally confounded both his doctors and his friends.

The unusual quality of Miss Evans' writing lies in the happy blend she has achieved of the good qualities of both artist and scientist. Like the former, she immerses herself in the material more deeply and for a longer period than the impatient scientist is generally willing to do. Long interviews with her subject alternate with patient study of all available documents and interviews with many different people who have known the subject. Unlike the scientist also, she seeks to avoid explanation, pointing to causes, and theories of personality. And, unlike the artist, she rigorously avoids the "creative" act of fashioning a character out of the many bits and pieces that may have come from as many different people. Instead she seeks to reflect into the mind's eye and ear of the reader an accurate reproduction, a "portrait," of what her subject actually looks like and how he sounds. She sets forth her picture, to a very considerable extent, in the very words the subject used in talking to her.

Those of us who have used these cases in our teaching know how absorbing the students find them, how eagerly they become concerned with all kinds of questions that come to mind, how profuse and varied are the hypotheses they bring forth to explain, to predict, to understand. And these products of the students' working with the material have the cardinal virtue of being *their own*, for Miss Evans offers none. Because she has saturated herself with the material, and because she brings to bear all her art and skill as a writer, she simply makes the person *live* in the perceptions of the reader.

Now, if we wish to inculcate the disciplines of

scientific caution and systematic verification, we find this material well adapted to such purposes. Once the hypotheses have been brought forth and noted down, we can go back to the document and test them against the facts. We may even invite the student to interrupt his reading at some point and undertake the hazardous venture of attempting a prediction or two. For example, most students who used the original Rocco material, which takes him to February 1947, predicted a far less well adjusted and socially productive Johnny than the follow-up material in this book, which brings him up to September 1952, shows him to have become.

The first two "portraits" in this book are of men at least a significant portion of whose lives was lived under conditions of real poverty. In the case of the third man, however, this was not the case. "Martin Beardson's" father was a prosperous business man, his mother an "upper class" woman of "genteel tradition." Her people, Martin says, "were the leading people in town." "Except for brief periods, and then only in recent years, Martin has never known deprivation in terms of food, clothing, and comfortable shelter. By realistic standards he also enjoyed many luxuries .... Yet Martin has always had a sense of poverty no less poignant than if he had been a slum child, and a sense of grievance no less biting than if he had lived all his life on the periphery of want" (p. 195). The student will ask "Why?" and the answer, I dare say, is not to be found solely in the fact that Martin's father and mother separated when he was but two years old.

An ever-present problem of scientific investigations that comes to mind again while reading these lives is that of the effect of the operation of observation on the very process observed. Miss Evans writes, in a footnote to "Miller," "It is to be borne in mind that the censure against which Miller constantly pleaded existed solely in his mind. Although I was genuinely interested and sympathetic, my primary relationship to my subjects is that of 'historian,' and not of counselor; in so far as humanly possible, my attitude was one of friendly neutrality" (p. 144). Even so, clinical experience has clearly demonstrated that the role the questioner assigns himself and the role in which he is perceived by the one he questions are frequently, if not always, different. And I found myself wondering, especially while reading the follow-up of Rocco, what effects, in defining his own roles and goals, the many hours of association with Miss Evans may have had on Johnny. That Miss Evans was not unaware of this possibility seems clearly indicated in the passage quoted above.

One wonders also why it is that Miss Evans herself, and I imagine many of her readers as well, comes to think of two of these men by their first names and one of them by his last. It is Johnny and Martin, but never William or Bill, just Miller.

This is a book that will be read with pleasure and pondered with profit by many scientists and professionals in clinical, social, and dynamic psychology, in sociology, psychotherapy, social work, and a number of other related fields. It will be of wide and lasting usefulness in classes and seminars in any field of study where the focus is on the lives of real people as they are actually lived in this world of ours.

Brooklyn College AUSTIN B. WOOD

THE ACHIEVEMENT MOTIVE. By David C. McClelland, John W. Atkinson, Russell A. Clark, & Edgar L. Lowell. New York: Appleton-Century-Crofts, 1953. Pp. xxii + 384. Price $6.00.

This book reports a swarm of little experiments, no one of which is more than a pilot study, but all of which are focused on one human need. The host of miniature demonstrations creates a fascinating pattern, from which the reader, like the authors, may learn "not only a lot about the achievement motive but other areas of personality as well." At the same time, the reader keeps glimpsing idea after idea for use in his own future research. He finds provocative inductive treatment of the nature of the achievement motive, its effects on behavior, especially school performance, and its origins in infancy and culture. All this "builds up the total picture out of many small experiments by a slow process of going from fact to hypothesis and back to fact again." It makes a methodologically fascinating illustration of what can be done inductively to explore a new field when the explorers possess the proper wit.

The experimental studies are preceded by a long theoretical chapter by McClelland on the nature of animal and human motives. In this chapter, as in the Introduction, one perceives that the author has come only a little way beyond the crude understanding of motivation provided by animal studies, in which the drives are always "primary" and arousal nearly always by deprivation. Even though at one point the writer explicitly acknowledges that human secondary needs do not have to be in a state of deficit to be aroused and that an earlier study of the projection of need Food may not be the best preparation for the study of the projection of need Achievement, the important differences between primary drives and secondary motives are never sufficiently explored. Even though McClelland acknowledges himself to be nearer to Hebb than to Hull, he starts from the reductionist postulate that "a motive is the learned result of pairing cues with affect," and so dooms himself to underestimate the sophistication of his human experimental animals. One is a bit startled to read that the *only* difference between his own motivation and that of a dog in heat

is that in him "behavior is controlled not so much by afferent stimulation but by anticipation." Drive-reduction postulates apparently do not lend themselves to anything like a realistic description of "the realm of value." It seems to this reviewer that both the theory propounded in McClelland's opening chapter and the measure that became the basis of all his experiments are therefore based on a dubious premise.

The consequence for experimental design was that "we accepted the notion from experimental studies of animals that motives could be experimentally aroused and their intensity controlled by manipulating arousal conditions. Our problem, then, boiled down to the attempt to arouse and control the intensity of a human motive and to measure its effect on imagination or fantasy. . . ." The projective hypothesis, derived from clinical psychology, was thus by-passed. However, the author only says, allusively, that "in this way we would not be forced to make any a priori assumptions as to what characteristics of fantasy indicate the presence and intensity of a certain motive." Rather, as McClelland has remarked elsewhere, the TAT signs of strong need Achievement could be identified "by a kind of content analysis" of stories told by subjects under control conditions and under conditions of reputedly strong laboratory arousal. The anomalous result is a 300-page book about the use of a modified Thematic Apperception Test in which the projective hypothesis is never discussed!

The new, laboratory-produced measure is called "need Achievement," we are told, "after Murray!" It bears little resemblance to that need as Murray has defined it. Nor do the authors seem to be aware of the confusion they have created. In their study of the infantile roots of need Achievement, for example, the experimental design tacitly assumes that their need Achievement, defined by the effect of verbally stimulating experimental subjects a few minutes before testing, and Murray's need Achievement, defined by evidence of an enduring orientation in the person, are identical.

Since, quite rightly, the authors anticipate that "the problems uncovered in the course of this research have been so numerous and complex that . . . their solution will require the cooperative effort of many scientists," one anticipates chaos! Many who will want to, or already have begun to, follow up on McClelland's leads will be accustomed to the more traditional use of the TAT. Worse, many who are only now becoming familiar with the TAT may uncritically apply McClelland's nonprojective measure when making clinical use of the projective test. But if confusion needs confounding, let us envisage what will happen when, as McClelland *et al.* anticipate, their "practical method of measuring one of the most important human motives . . . can be applied to other motives with equal success!"

Let us hope that before we have two of every human need, occasion may be found by some experimenter to discover the empirical results of measuring the two need Achievements side by side.

CHARLES C. MCARTHUR
Department of Hygiene
Harvard University

WHO SHALL SURVIVE? FOUNDATIONS OF SOCIOMETRY, GROUP PSYCHOTHERAPY AND SOCIODRAMA. (Rev. Ed.) By J. L. Moreno. Beacon, New York: Beacon House, 1953. Price $10.00.

GROUP DYNAMICS: RESEARCH AND THEORY. Edited by Dorwin Cartwright and Alvin Zander. Evanston, Ill.: Row, Peterson & Co., 1953. Price $6.00.

In the Foreword to his book, *Personality: a Biosocial Approach to Origins and Structure*, Gardner Murphy described the volume as somewhat like "an explorer's kit, containing, to be sure, some standard tools, and also some maps. Some of the maps are sober, some eccentric, no doubt; but all are drawn in the belief that any map of a far country encourages more travel than an architect's finished representation of the doorway as it is here and now" (p. x). Imbued with this philosophy, the present reviewer welcomes the appearance of the revised edition of *Who Shall Survive?* and of *Group Dynamics*. Each is an important volume which has much to contribute to improving our understanding of the social nature of man and the nature of man's social behavior. To those of us searching for a "world" view in the field of social psychology, the special projections of the cartography of these volumes must be properly identified before the maps can be redrawn and integrated. Moreno offers us a complete theoretical globe. "Sociometry cuts through all social sciences as it deals with social phenomena at a deep level where they merge or more precisely before they 'e'merge into 'psychological,' 'sociological,' 'anthropological,' or 'economic' phenomena" (p. 54). Cartwright and Zander, on the other hand, offer us half a dozen illustrations of the cartographer's art without too clear an indication of how they are interrelated, what the contiguous areas may be like, on what basis they were chosen—but each does depict a large and rich terrain, and the maps are beautiful.

Though these two books are very different in many ways, it seems desirable to consider them in a single review in order to contribute a counter, however feeble, to the schismogenic forces currently operative within the field. Because each is likely to appeal to somewhat different audiences, on temperamental grounds, it is well to reaffirm

the need for cognitive awareness of their relatedness. Perhaps it would be well to note some of the differences that are likely to evoke differential conative responses—in the hope that forewarned is forearmed. Moreno's writing includes a large measure of personal testament; Cartwright and Zander's work is a collection of 35 papers to which the editors have added six integrative theoretical "introductions"; only one of the reprinted papers has an editor as author. Cartwright and Zander are very modest in their offerings. Moreno rejects the desirability of modesty. Cartwright and Zander conform to the sobriety of the "scientific" style of writing; Moreno is, in much of the book, spontaneous, personal, anecdotal. The Moreno volume is much longer than the Cartwright and Zander book; it contains over 200 more pages, stands a full inch wider on the shelf (though of the same height and depth), and is correspondingly heavier. Let us turn now, however, to a consideration of each volume by itself.

Dr. Moreno's new version of *Who Shall Survive?* retains the same central research material that was presented in the 1934 original. However, the context is very different. The development of the sociometric movement during the intervening 20 years is reviewed in "Preludes," which includes also additional material on the pre-1933 background. It is this section of about 100 pages that departs most from traditional scientific writing. However, this is Moreno's statement of the history of "his" approach. The pages are given Roman numerals, and even if unusually long, are completely in the character of prefaces to which we are more accustomed. There are in the Moreno personal statement such items as characterization of the book as a bible ("The Words of the Father," the old testament; "Who Shall Survive?" the new, p. lxvii); an insistence on sole authorship of the book, emphasizing that "An idea book like WHO SHALL SURVIVE? cannot be conceived 'in collaboration'" (p. lxxxvii) (yet in the 1934 edition, Helen H. Jennings is described as a collaborator); the description of an imaginary correspondence with H. A. Murray in which "he sends me letters which he never actually wrote and I send him letters, often in the middle of the night, which he receives promptly and without benefit of mail carriers" (p. xc); and an attack upon the integrity of some of the leaders of the current group dynamics movement (p. cii ff.). However future historians may accept or "correct" the version Moreno gives, these Preludes are important documentation for a history of the social science, and valuable for a study of the process of writing history. Moreno recognizes that there is far more controversy about his person than there is about his theories. He concludes this presentation with the statement: "I have tried to explain why an individual of my type was particularly suited to produce them. I hope that I succeeded some and that everyone, friends and enemies, will enjoy them, because they are autobiography 'unadulterated.' Now I fear only that many will read the Preludes and not the book" (p. cviii). Your reviewer shares these fears and has dwelt on this section of the book to make explicit what is likely to prove the major drawback for many readers. Other readers, however, will appreciate the insights this kind of statement will give them into the nature of the originator of one of the most productive approaches currently available to students of social processes.

The body of the book remains much as in the original. The first five parts, the sociometric system, evolution of groups, sociometry of a community, construction and reconstruction of a community, and sociometric planning of society, follow, as in the first edition, the report of original researches, with only minor changes in the discussion, paragraphing, and headings. The final section, however, "Who Shall Survive?" is greatly elaborated. Instead of the illustrative materials included in the 1934 version, we now have about 100 pages of "The Sociometric System and Advanced Sociometric Theory." In this section, in addition to summaries of further research data, about 30 pages are devoted to developing the statistics for sociometric measurement and research based on deviation of obtained from chance findings. Finally, starting on page 696, there are 20 pages of "General Hypotheses and Recommendations for Further Research." In some ways these hypotheses may be considered as introduction, in some ways as summary for the 1953 position of sociometry.

It is interesting to note that the first edition contained only 41 items on a one-page bibliography, while the new volume claims over 1,300 references in a 31-page bibliography. This single quantitative index may serve as a clue to the impact of Moreno's ideas during the 20-year period, and the difference between the original and the second edition of the book.

The volume edited by Cartwright and Zander brings us to the first major effort to organize and make available the theory and findings, the methods and interpretations of those identified with the "group dynamics" movement. The editors felt that it would be premature to prepare a systematic "text" inasmuch as there were still too many areas of thinking for which there was no empirical test yet, and too many significant studies that would not fit easily into a single theory. They have chosen to combine integrative interpretation, in the form of introductions to each of the six sections of the book, with the reproduction of the original research papers that have contributed most significantly to their formulations. It is, therefore, both text and readings. The introductions provide general perspective on the selected study areas, overviews and integrations of the

readings included, and questions for further research.

The book is divided into six parts: approaches to the study of groups, group cohesiveness, group pressures and group standards, group goals and group locomotion, the structural properties of groups, and leadership. Each part stands by itself, with introduction and from three to ten readings. Though the introductions establish the theoretical orientation toward the group dynamics approach (and of this approach to various problems), there is no clear continuity from one part to another. Thus, part five, on the structural properties of groups, does not refer to leadership, and part six, on leadership, seems independent of the discussion of the structural properties of groups.

The 35 readings range in time of original appearance from one by French which appeared in 1941 through one specially prepared by Kahn and Katz for this volume to two still to appear in other vehicles (but already committed). Thirty may be identified as having appeared in the 1950's. Ten of the selections are from *Human Relations*, eight from the *Journal of Abnormal and Social Psychology*, six are selections from books, three from the *American Sociological Review*, and the remainder single borrowings from various other sources.

The editors describe their criteria for the selection of readings thus: "With few exceptions, we have included only reports that present empirical findings based upon quantitative research, and when several related articles were available, we attempted to choose the one which best summarized the total set of research. Some of the classics have, accordingly, been left out because more recent publications summarize them as well as provide additional findings" (p. xi). This statement of intention is well borne out by the contents. They well reflect the vitality of the research programs that may be identified with "group dynamics."

The editors also suggest that "a reader who desires a brief overview of the field of group dynamics will discover that a consecutive reading of the introductory chapters is useful for this purpose" (p. xi). This is probably true, but the present reviewer feels that within this framework more could have been done. The first part of the book is primarily theoretical and tries to describe some of the major varieties of approaches to the study of groups. It includes selections by Cattell, Bales, Stogdill, Scheidlinger, and Jennings. By no means would all of these be identified with what is usually considered "group dynamics" research, nor are they a complete review of the approaches currently employed. On a sampling basis, they may serve some purpose in indicating the variety of approaches being made to similar problems. Yet in the Preface, the editors acknowledge that they had to omit work that is more clearly within the framework of the group dynamicists, for example, the work of Herbert Thelen, William F. Whyte, and others. To the extent that one is interested in the variety of research approaches to these problems, some of the nonlaboratory approaches ought to have been included. To the extent that one is really concerned with an overview of group dynamics, some of the stimulating special emphases and new directions must be added to the reading list.

Nevertheless, this book does serve a very useful function in assembling much of the major work in the major areas in which the study of group dynamics has progressed, and in providing the six general theoretical statements and integrations, with the suggestions for further work.

Developments within the field of social psychology have been rapid. Theory, research, and application have all developed apace. It is increasingly difficult to maintain a general perspective and to "keep up" with the current output. The two books here discussed illustrate the difficulties. Each represents a "line" one may pursue with great profit, yet the overlap in terminology, research methods, reference groups, and training, and even the channels of communication used, is relatively slight. And there are other lines that are nowhere represented. What of the work of Trigant Burrow and the development of phyloanalysis? What of the promising trends in group psychotherapy outside of the sociometric pattern? What of the continued developments in attitude, opinion, and communication research? These are but a few of the major trends that are readily identified. The alert specialist must go farther. What new insights and points of departure does the transactional approach offer for the study of social behavior? As each line of promise becomes more fully worked out, in the light of the reasonableness of the approach and the significance of the findings, it becomes increasingly difficult to establish the relationships between figure and ground in the perception of the general field and to guide such organization rather than to let the stimulus and autochthonous factors prevail.

EUGENE L. HARTLEY
The City College, New York

NEW BOOKS RECEIVED

ALLPORT, G. W. *The nature of prejudice.* Cambridge, Mass.: Addison-Wesley Publishing Co., 1954. Pp. xviii + 537. Price $5.50.

BACH, G. R. *Intensive group psychotherapy.* New York: Ronald, 1954. Pp. xi + 446. Price $6.00.

BARRON, M. L. *The juvenile in delinquent society.* New York: Knopf, 1954. Pp. xix + 349 + vii. Price $5.00.

BAUCHARD, P. *The child audience: a report on press, film and radio for children.* Paris: Unesco, 1952. Pp. 198. Price $2.00.

BERGLER, E., & KROGER, W. S. *Kinsey's myth of female sexuality.* New York: Grune & Stratton, 1954. Pp. viii + 200. Price $3.75.

BETTELHEIM, B. *Symbolic wounds.* Glencoe, Ill.: Free Press, 1954. Pp. 286. Price $4.75.

BLACK, P. *The mystique of modern monarchy.* London: C. A. Watts & Co., 1953. Pp. xii + 92. Price 8s 6d.

BLACKHURST, J. H. *Body-mind and creativity.* New York: Philosophical Library, 1954. Pp. 186. Price $3.00.

BLAKE, FLORENCE G. *The child, his parents and the nurse.* Philadelphia: Lippincott, 1954. Pp. xvii + 440. Price $5.00.

BURROW, T. *Science and man's behavior.* New York: Philosophical Library, 1953. Pp. xii + 564. Price $6.00.

CHRISTIE, R., & JAHODA, MARIE. (Eds.) *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954. Pp. 279. Price $4.50.

DREIZEN, L. *Neurosis, crime and progress.* Mesnil, France: Typographie Firmin-Didot & Cie, 1953. Pp. 88. No price supplied.

DREIZEN, L. *Shakespeare, Marc Antony and mass persuasion.* Mesnil, France: Typographie Firmin-Didot & Cie, 1953. Pp. vii + 67. No price supplied.

DURKHEIM, E. *Sociology and philosophy.* (Trans. by D. F. Pocock.) Glencoe, Ill.: Free Press, 1953. Pp. xli + 97. Price $2.50.

ELLIS, A. *The American sexual tragedy.* New York: Twayne Publishers, 1954. Pp. 288. Price $4.50.

EVANS, JEAN. *Three men. An experiment in the biography of emotion.* New York: Knopf, 1954. Pp. xviii + 297. Price $3.75.

EYSENCK, H. J. *Uses and abuses of psychology.* Baltimore: Penguin Books, 1953. Pp. 318. Price $ .65.

FRANK, L. K. *Feelings and emotions.* New York: Doubleday, 1954. Pp. vi + 38. Price $ .85.

FRENCH, T. M. *The integration of behavior.* Vol. II: *The integrative process in dreams.* Chicago: Univer. of Chicago Press, 1954. Pp. xi + 367. Price $6.50.

FREUD, S. *On aphasia.* (Trans. by E. Stengel.) New York: International Universities Press, 1953. Pp. xv + 105. Price $3.00.

FRIEDENBERG, E. Z., & ROTH, J. A. *Self-perception in the university: a study of successful and unsuccessful graduate students.* Chicago: Univer. of Chicago Press, 1953. Pp. vii + 102. Price $1.75.

GANNON, T. *Psychology: the unity of human behavior.* Boston: Ginn & Co., 1954. Pp. xii + 482. Price $4.75.

GANZ, MADELAINE. *The psychology of Alfred Adler and the development of the child.* New York: Humanities Press, 1953. Pp. xxiii + 203. Price $4.50.

GARDNER, G. E. (Ed.) *Case studies in childhood emotional disabilities.* Vol. I. New York: Amer. Orthopsychiat. Ass., 1953. Pp. vii + 368. Price $5.00.

GEHMAN, R. *Driven.* New York: David McKay Co., 1954. Pp. 307. Price $3.50.

GILLIN, J. (Ed.) *For a science of social man.* New York: Macmillan, 1954. Pp. vii + 289. Price $4.00.

GLADWIN, T., & SARASON, S. B. *Truk: man in paradise.* New York: Wenner-Gren Foundation for Anthropological Research, 1954. Pp. 655. No price supplied.

GOULDNER, A. W. *Patterns of industrial bureaucracy.* Glencoe, Ill.: Free Press, 1954. Pp. 282. Price $4.50.

GRAY, J. S. *Psychology applied to human affairs.* (2nd Ed.) New York: McGraw-Hill, 1954. Pp. vii + 581. Price $6.00.

HARDING, D. W. *Social psychology and individual values.* New York: Longmans, Green, 1953. Pp. vii + 184. Price $2.40.

INGHAM, K. V., & LOVE, LEONORE R. *The process of psychotherapy.* New York: McGraw-Hill, 1954. Pp. ix + 270. Price $5.00.

KAMIAT, A. H. *The ethics of civilization.* Washington: Public Affairs Press, 1954. Pp. vi + 80. Price $2.00.

KARPMAN, B. *The sexual offender and his offenses.* New York: Julian Press, 1954. Pp. xiii + 744. Price $10.00.

LANDER, B. *Towards an understanding of juvenile delinquency.* New York: Columbia Univer. Press, 1954. Pp. xv + 143. Price $3.00.

LIBO, L. M. *Measuring group cohesiveness.* Ann Arbor: Univer. of Michigan Press, 1954. Pp. ix + 111. Price $2.00.

LLOYD, FRANCES. *Educating the sub-normal child.* New York: Philosophical Library, 1954. Pp. vii + 148. Price $3.75.

LOOMIS, C. P., *et al. Rural social systems and adult education.* East Lansing, Mich.: Michigan State Coll. Press, 1953. Pp. viii + 392. Price $5.00.

MCCURDY, H. G. *The personality of Shakespeare.* New Haven: Yale Univer. Press, 1953. Pp. xi + 243. Price $5.00.

MCGARR, MARGARET A. *And lo, the star.* New York: Pageant Press, 1953. Pp. 116. Price $2.50.

MCLEOD, L. S. *Mental health in the home.* New York: Bookman Associates, 1953. Pp. 243. Price $3.50.

MOLONEY, J. C. *Understanding the Japanese mind.* New York: Philosophical Library, 1954. Pp. xviii + 252. Price $3.50.

NUTTIN, J. *Tâche réussite et échec: théorie de la conduit humaine.* Antwerp, Belgium: N. V. Standaard-Boekhandel, 1953. Pp. ix + 530. No price supplied.

PERRY, R. B. *Realms of value.* Cambridge, Mass.: Harvard Univer. Press, 1954. Pp. xii + 497. Price $7.50.

ROSE, A. M. *Theory and method in the social sciences.* Minneapolis: Univer. of Minnesota Press, 1954. Pp. xii + 351. Price $5.00.

ROTHNEY, J. W. M. *The high school student: a book of cases.* New York: Dryden Press, 1953. Pp. xv + 271. Price $1.90.

SAENGER, G. *The social psychology of prejudice.* New York: Harper, 1953. Pp. xv + 304. Price $4.00.

SAPPENFIELD, B. R. *Personality dynamics: an integrative psychology of adjustment.* New York: Knopf, 1954. Pp. xiv + 412 + xvi. Price $5.50.

SCHILDER, P. *Medical psychology.* (Trans. and ed. by D. Rapaport.) New York: International Universities Press, 1953. Pp. 428. Price $7.50.

SHAPIRO, H. L. *Race mixture.* Paris: Unesco, 1954. Pp. 56. No price supplied.

SIMPSON, G. *Man in society.* New York: Doubleday, 1954. Pp. vi + 90. Price $ .95.

SMITH, A. J. *Immortality.* New York: Prentice-Hall, 1954. Pp. 248. Price $3.00.

SMOLLAR, L. *Know your doctor.* Boston: Little, Brown, 1953. Pp. 173. Price $3.00.

SPELLMAN, H. H. *Successful management of matrimonial cases.* New York: Prentice-Hall, 1954. Pp. xiv + 306. Price $5.65.

STOLZ, LOIS M., AND COLLABORATORS. *Father relations of war-born children.* Stanford: Stanford Univer. Press, 1954. Pp. viii + 365. Price $4.00, paper.

STOTT, D. H. *Saving children from delinquency.* New York: Philosophical Library, 1953. Pp. x + 266. Price $4.75.

TAYLOR, W. S. *Dynamic and abnormal psychology.* New York: American Book Co., 1954. Pp. xiv + 658. Price $5.50.

VEDDER, C. B. *The juvenile offender.* New York: Doubleday, 1954. Pp. xii + 510. Price $6.00.

WEINBERG, S. K. *Society and personality disorders.* New York: Prentice-Hall, 1952. Pp. viii + 536. Price $7.65.

WITKIN, H. A., *et al. Personality through perception.* New York: Harper, 1954. Pp. xvi + 571. Price $7.50.

WOLFFHEIM, NELLY. *Psychology in the nursery school.* (Trans. by C. L. Hannam.) New York: Philosophical Library, 1953. Pp. 144. Price $3.75.

WRIGHT, A. F. (Ed.) *Studies in Chinese thought.* Chicago: Univer. of Chicago Press, 1953. Pp. xiii + 317. Price $4.00.

YOUNG, K. *Isn't one wife enough?* New York: Henry Holt, 1954. Pp. xiv + 476. Price $6.00.

Studi della facolta medica senese. Siena, Italy: Atti Accademia dei Fisiocritici, Siena, Sezione Medico-Fisica, 1953. Pp. 516. No price supplied.

Neurology and Psychiatry

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles on neurological survey in addition to general clinical neurology and experimental neurophysiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition the following special branches are dealt with fully: psychoanalysis, individual and analytical psychology; serology, criminology, alcoholism and drug addiction as related to psychiatry; psychosomatic medicine; social and industrial psychology and psychiatry, vocational guidance and mental testing; psychometrics; personality testing and the Rorschach test; heredity and statistical studies as they apply to psychiatry. The subjects of mental defect and epilepsy are fully reported.

Annual volume of 1000 pp., including an index classified by author and subject
Subscription price $22.50 per annual volume

EXCERPTA MEDICA SERVICE CORPORATION

250 Madison Avenue — New York 16, N. Y.

Journal of Abnormal and Social Psychology

AVAILABLE BACK ISSUES

July 1, 1951

Year	Volume	Available Numbers				Price per Number	Price per Volume
1906-07	1	—	—	—	—	—	—
1907-08	2	—	—	—	5	$1.00	—
1908-09	3	—	—	—	—	—	—
1909-10	4	—	—	—	—	—	—
1910-11	5	—	2	—	—	$1.00	—
1911-12	6	—	—	—	—	—	—
1912-13	7	—	—	—	—	—	—
1913-14	8	—	—	—	—	—	—
1914-15	9	—	—	—	—	—	—
1915-16	10	—	—	—	—	—	—
1916-17	11	—	—	—	—	—	—
1917-18	12	—	—	—	—	—	—
1918-19	13	3	5	—	—	$1.00	—
1919-20	14	—	5	4	5	$1.00	—
1920-21	15	3	—	5	4	$1.00	—
1921-22	16	—	—	4	—	—	—

		Apr.	July	Oct.	Jan.		
1922-23	17	3	—	—	—	$0.50	—
1923-24	18	1	3	3	1	$0.50	$5.00
1924-25	19	—	—	—	—	—	—
1925-26	20	—	—	—	—	—	—
1926-27	21	—	—	—	—	—	—
1927-28	22	—	—	1	1	$0.50	—
1928-29	23	—	1	1	1	$0.50	—
1929-30	24	1	1	1	1	$0.50	$5.00
1930-31	25	1	1	1	1	$0.50	$5.00
1931-32	26	1	1	1	1	$0.50	$5.00
1932-33	27	1	1	1	1	$0.50	$5.00
1933-34	28	1	1	1	1	$0.50	$5.00
1934-35	29	1	1	1	1	$0.50	$5.00
1935-36	30	1	1	1	1	$0.50	$5.00
1936-37	31	1	1	1	1	$0.50	$5.00

		Jan.	Apr.	July	Oct.		
1937	32	1	1	1	1	$0.50	$5.00
1938	33	1	1	1	1	$0.50	$5.00
1939	34	—	—	—	—	—	—
1940	35	1	1	1	1	$0.50	$5.00
1941	36	1	1	1	1	$0.50	$5.00
1942	37	1	1	1	1	$0.50	$5.00
1943	38	1	1	1	1	$0.50	$5.00
1944	39	1	1	1	1	$0.50	$5.00
1945	40	1	1	1	1	$0.50	$5.00
1946	41	1	1	1	1	$0.50	$5.00
1947	42	1	1	1	1	$0.50	$5.00
1948	43	1	1	1	1	$0.50	$5.00
1949	44	1	1	1	1	$0.50	$5.00
1950	45	1	1	1	1	$0.50	$5.00
1951	46	By subscription 1/46				$2.50	—

Postage prepaid on U. S. orders. Add $.50 per volume on foreign orders. All stock subject to prior sale.

The American Psychological Association gives the following discounts:

10% on orders of $50.00 or over for any one journal
20% on orders of $100.00 or over for any one journal
30% on orders of $150.00 or over for any one journal

Orders for available back issues and for current subscriptions should be addressed to:

American Psychological Association

1333 Sixteenth Street N. W. Washington 6, D. C.